



# **Ansiktsigenkänning, röst- och rörelsestyrning i vardagsbruk**

En utredning med Kinect och Samsungs Smart TV i fokus

Jeanette Weckström

Examensarbete  
Informations- och medieteknik  
2013

EXAMENSARBETE	
Arcada	
Utbildningsprogram:	Informations- och medieteknik
Identifikationsnummer:	4189
Författare:	Jeanette Weckström
Arbetets namn:	Ansiktsigenkänning, röst- och rörelsestyrning i vardagsbruk – En utredning med Kinect och Smart TV i fokus
Handledare (Arcada):	Johnny Biström
Uppdragsgivare:	
<p>Det blir allt vanligare med apparater som hanterar ansiktsigenkänning, röst- och rörelsestyrning. Ett användargränssnitt med den sortens funktioner kallas för ett naturligt användargränssnitt, eftersom interaktionsmetoderna anses vara naturliga för en människa. Det har redan länge forskats i ansiktsigenkänning, röst- och rörelsestyrning, men ämnet är dock aktuellt eftersom produkter med sådana funktioner först nu märkbart har börjat dyka upp på marknaden. Hur välutvecklade är dessa funktioner egentligen och hurdan teknik ligger bakom dem? Syftet med detta arbete är att undersöka hur användbara naturliga användargränssnitt är jämfört med dagens fjärrkontroller. Forskningen fokuseras kring två vardagliga produkter som innehåller dessa funktioner: Microsofts Kinect och Samsungs Smart TV. Arbetet behandlar endast kontroll- och beröringsfri rörelsestyrning. Undersökningsobjektens andra egenskaper och hur de är som produkter, tas inte heller ställning till. Arbetet innehåller en teoridel där begreppen naturligt användargränssnitt, ansiktsigenkänning, röst- och rörelsestyrning behandlas. Därefter går det djupare in på tekniken bakom Kinect för att få en bild av hur funktionerna i fråga är uppbyggda i en konkret produkt. Slutligen så presenteras Smart TV:ns naturliga användargränssnitt. Efter teoridelen följer en testdel där ansiktsigenkänningen, röst- och rörelsestyrningen i Kinect och Smart TV:n testas. Testerna bevisade att knappar och fysisk kontakt fortfarande behövs. I dagens läge finns det sådana ansiktsigenkänningsfunktioner, röststyrningsfunktioner och rörelsestyrningsfunktioner som är riktigt användbara, men de fungerar då bra tillsammans med de klassiska interaktionsmetoderna, inte ensamma. Det beror också på produkten, om det är någon idé med sådana interaktionsmetoder, eller inte. Då funktionerna slås ihop med rätt produkt, kan de faktiskt vara till nytta. De bör dock vara flexibla och snabba samt enkla att använda för att man inte skall tappa intresset för att använda dem. Om de inte är det så kommer man antagligen att övergå till att använda fjärrkontrollen även om funktionerna verkar spännande i början.</p>	
Nyckelord:	Ansiktsigenkänning, röststyrning, rörelsestyrning, Kinect, Smart TV, naturliga användargränssnitt, interaktion
Sidantal:	61
Språk:	Svenska
Datum för godkännande:	24.4.2013

DEGREE THESIS	
Arcada	
Degree Programme:	Information- and Media Technology
Identification number:	4189
Author:	Jeanette Weckström
Title:	Face recognition, voice and motion control in everyday use – An inquiry with Kinect and Samsung's Smart TV in focus
Supervisor (Arcada):	Johnny Biström
Commissioned by:	
<p>Devices that handle face recognition, voice and motion control are becoming increasingly common. A user interface with that kind of features is called a natural user interface, since the interaction methods are regarded as natural for a human. People have already been researching in face recognition, voice and motion control for a long time, but the subject however, is topical since products with such features only now noticeably have begun to appear on the market. How well developed are these features really and what kind of technology lies behind them? The purpose of this thesis is to investigate how useful natural user interfaces are compared to today's remotes. The research is focused around two everyday products that contain these features: Microsoft's Kinect and Samsung's Smart TV. The thesis only deals with controller-free and non-contact motion control. Other properties of the researched products and how they are as products, is not investigated. The thesis contains a theoretical part where the concepts natural user interface, face recognition, voice and motion control are discussed. Next, the technology behind Kinect is discussed deeper to get a picture of how the functions in question are built in a physical product. Finally, the natural user interface of the Smart TV is presented. The theoretical part is followed by a test part where the face recognition, voice and motion control in Kinect and Smart TV are tested. The tests proved that buttons and physical contact is still needed. There are presently such facial recognition functions, voice control functions and motion control functions that are really useful, but they work well together with the classical interaction methods, not alone. It also depends on the product, whether there is a point in having such interaction methods or not. When the functions are combined with the right product, they can actually be useful. They should however be flexible, quick and easy to use so that one does not lose interest in using them. If they are not like that, one will probably start using the remote control again, even if the functions seems exciting at first.</p>	
Keywords:	Face recognition, voice control, motion control, Kinect, Smart TV, natural user interfaces, interaction
Number of pages:	61
Language:	Swedish
Date of acceptance:	24.4.2013

# INNEHÅLL

<b>1</b>	<b>INLEDNING .....</b>	<b>8</b>
1.1	Bakgrund .....	9
1.2	Syfte och mål .....	9
1.3	Avgränsning .....	10
1.4	Metoder .....	10
1.5	Terminologi och förkortningar .....	11
<b>2</b>	<b>NATURLIG INTERAKTION .....</b>	<b>12</b>
2.1	Rörelsestyrning .....	13
2.2	Röststyrning .....	14
2.2.1	<i>Visuell taligenkänning .....</i>	<i>15</i>
2.3	Ansiktsgenkänning .....	15
2.3.1	<i>Ansiktsgenkänning på ett längre avstånd .....</i>	<i>16</i>
<b>3</b>	<b>KINECT .....</b>	<b>17</b>
3.1	Komponenterna .....	17
3.2	Hur djupmätningen fungerar .....	18
3.2.1	<i>IR-laserprojektorns egenskaper .....</i>	<i>19</i>
3.2.2	<i>Störningar från omgivande ljus .....</i>	<i>20</i>
3.2.3	<i>RGBD .....</i>	<i>21</i>
3.3	Skeletal-tracking och hur rörelsestyrningen fungerar .....	21
3.3.1	<i>Kinect tolkar varje led skilt för sig .....</i>	<i>22</i>
3.3.2	<i>Objektigenkänning och per-pixel klassificering .....</i>	<i>22</i>
3.3.3	<i>Maskininlärning och träningsdata .....</i>	<i>23</i>
3.3.4	<i>Varför basera rörelsestyrningen på djupmätning? .....</i>	<i>24</i>
3.3.5	<i>Hur man använder rörelsestyrningen i Kinect .....</i>	<i>24</i>
3.4	Hur röststyrningen fungerar .....	25
3.4.1	<i>Ljudbehandling .....</i>	<i>26</i>
3.5	Kinect Identity .....	26
3.5.1	<i>Ansiktsgenkänning .....</i>	<i>27</i>
3.5.2	<i>Personidentifiering .....</i>	<i>28</i>
<b>4</b>	<b>SMART TV .....</b>	<b>30</b>
4.1	Rörelsestyrning .....	31
4.1.1	<i>Hur används rörelsestyrningsfunktionen? .....</i>	<i>31</i>
4.1.2	<i>Ljuset och bakgrundens inverkan på rörelsestyrningen .....</i>	<i>32</i>
4.2	Röststyrning .....	32
4.3	Ansiktsgenkänning .....	34

<b>5</b>	<b>TESTER OCH RESULTATREDOVISNING .....</b>	<b>34</b>
5.1	Rörelsestyrningsfunktionen i allmänhet.....	36
5.1.1	<i>Menynavigering med rörelser .....</i>	<i>37</i>
5.1.2	<i>Rörelsestyrningens responstid .....</i>	<i>38</i>
5.1.3	<i>Rörelsestyrningens påfrestning.....</i>	<i>39</i>
5.2	Röststyrningen i allmänhet .....	40
5.2.1	<i>Röststyrning och omgivningens ljudnivå .....</i>	<i>41</i>
5.2.2	<i>Röststyrningens responstid .....</i>	<i>43</i>
5.2.3	<i>Röststyrning och uttal .....</i>	<i>44</i>
5.3	Ansiktsigenkänningen i allmänhet .....	44
5.3.1	<i>Hur omgivningens belysning påverkar ansiktsigenkänningen .....</i>	<i>45</i>
5.3.2	<i>Ansiktsigenkänningsfunktionens flexibilitet .....</i>	<i>46</i>
5.3.3	<i>Ansiktsigenkänning och personer med olika hudfärg.....</i>	<i>51</i>
5.4	Resultatredovisning .....	53
<b>6</b>	<b>SLUTSATSER OCH DISKUSSION .....</b>	<b>54</b>
	<b>Källor .....</b>	<b>56</b>
	<b>Bilagor .....</b>	<b>60</b>

## Figurer

Figur 1. Röststyrning handlar om att jämföra mönster (Rankin 2007). .....	15
Figur 2. Hur avstånd kan representeras genom användning av filter (Ban et al. 2008 s. 2410). .....	16
Figur 3. Hur Kinect ser ut utanpå (Kramer et al. 2012 s. 2). .....	18
Figur 4. a) IR-prickar kan ses i den vänstra bilden som är en närbild av området i den röda rutan till höger b) djupbild (Zhang 2012 s. 5.) .....	19
Figur 5. IR-laserns prickmönster där de nio starkare punkterna går att urskilja (Kramer et al. 2012 s. 12). .....	20
Figur 6. OpenNI kan bl.a. känna igen en hand, bearbeta skelett och känna igen rörelser (Cruz et al. 2012 s. 42). .....	21
Figur 7. Skeletal-tracking i Kinect. Djupmappens gråa nyanser som representerar avstånd färgas med gröna toner, eftersom grönt förknippas med Xbox. (Michel 2011).22	
Figur 8. En djupbilds pixlar färgläggs i olika färger beroende på vilken kroppsdel varje enskild pixel tillhör. Därefter kan Kinect räkna ut XYZ-koordinaterna för kroppens leder (Shotton et al. 2011 s 1297). .....	23
Figur 9. En förloppsindikator visar hur länge man skall hålla handikonen vid ett objekt för att välja det (Mejia 2010). .....	25
Figur 10. Hur Kinect ser ett ansikte. a) vanlig RGB-bild b) djupbild c) ansiktets struktur (Zhang 2012 s. 8). .....	27
Figur 11. Kinects ansiktsgenkänningsfunktion tar olika delar av ansiktet i beaktande skilt för sig (Cao et al. 2010 s. 2712). .....	28
Figur 12. Samsungs Smart TV ES8005 (Benzon 2012). .....	31
Figur 13. Då man uttalat röststyrningens aktiveringsfras dyker möjliga kommandon upp på skärmen (van Ballegoie 2012). .....	33
Figur 14. Ansiktsgenkänningsfunktionen sparar en miniatyrbild av en användares ansikte. ....	34
Figur 15. Mätinstrumenten som användes i samband med testerna. ....	35
Figur 16. Ansiktsbehåring och olika hårstilar. Första delen av testet och dess delmoment. ....	47
Figur 17. Ansiktsbehåring och olika hårstilar. Andra delen av testet och dess delmoment (hästsvans saknas). ....	48

Figur 18. Användning av solglasögon. Hur klarar ansiktsgenkänningsfunktionen av a) mörka solglasögon b) ljusa solglasögon med mindre lins? .....	49
Figur 19. Användning av glasögon eller huvudbonad. Hur klarar ansiktsgenkänningsfunktionen av a) glasögon b) en hatt? .....	49
Figur 20. Man testade hur smink påverkar Kinects och Smart TV:ns ansiktsgenkänningsfunktioner. ....	50
Figur 21. Testpersonen blev maskerad som mörkhyad. ....	52

## Tabeller

Tabell 1. Sanningsvärdetabell. Kinect kommer här att anta att identitet nr 1 är den rätta eftersom den inte har negativa värden och minst ett positivt värde. (Leyvand et al. 2011 s. 95) .....	29
Tabell 2. Ansiktsgenkänningens flexibilitet. ....	60

# 1 INLEDNING

Det blir allt vanligare med produkter som klarar av ansiktsigenkänning, röst- och rörelsestyrning, d.v.s. apparater som styrs med hjälp av ett s.k. naturligt användargränssnitt. Sådana funktioner har existerat redan länge, men det är nu som tekniken har börjat utvecklas på allvar.

Detta examensarbete innehåller en utredning över naturliga användargränssnitt i vardagslivet med fokus på Microsofts Kinect och Samsungs Smart TV (modell ES8005). Båda två innehåller alla tre funktioner som kommer att behandlas: ansiktsigenkänning, röststyrning och kontrollerfri rörelsestyrning. Jag kommer att undersöka hur naturliga användargränssnitt fungerar och om de är användbara. Arbetet innehåller först en förklaring på vad ett naturligt användargränssnitt är och därefter finns en allmän översikt på ansiktsigenkänning, röst- och rörelsestyrning. Eftersom Kinects funktionsprinciper till en stor del har avslöjats, så kommer jag därpå att ta en noggrannare titt på dess naturliga användargränssnitt. På det sättet fås både en allmän syn på ansiktsigenkänning, röst- och rörelsestyrning samt en utredning över hur dessa funktioner är uppbyggda i en konkret och dessutom välkänd produkt. Jag har varit i kontakt med Samsung flera gånger i hopp om att få någon information om det naturliga användargränssnittet i deras Smart TV, men de kunde inte avslöja någonting. Detta arbete går av den orsaken inte djupare in på själva tekniken i TV:n. Det slutliga svaret jag fick av Samsung lät så här: "Everything is more or less confidential in our world (tech documentation). This means we as employees don't even get it. I can only refer to the E-manual (which can be downloaded from our website). This is the documentation we as technical Engineers have". Eftersom produkter som styrs med ett naturligt användargränssnitt så nyligen har börjat dyka upp märkbart på marknaden, så funderar jag hur välutvecklade dessa egenskaper egentligen är. I min forskning kommer jag därför att kontrollera hur bra ansiktsigenkänningen, röst- och rörelsestyrningen i Kinect och Samsungs Smart TV egentligen fungerar och vilka eventuella brister de har.



## 1.1 Bakgrund

Att kunna styra apparater med den egna kroppen, var sådant som jag tidigare bara hade sett på film och då funderat om en sådan vardag någon gång skulle bli verklighet. Nu då sådana maskiner är ett faktum så är jag intresserad av att ta reda på hur de fungerar.

Då jag första gången hörde om Kinect och att man med denna apparat kan styra videospel helt och hållet med kroppen, så blev jag väldigt nyfiken. Jag var dock tvivelaktig över hur bra ett sådant användargränssnitt egentligen fungerar. Det att Kinect förutom rörelsestyrning även innehöll röststyrning och ansiktsgenkänning gjorde inte produkten mindre intressant. Samsung lanserade i fjol (2012) en smart TV med alla dessa tre funktioner. Då jag plötsligt hade tillgång till en sådan TV, och då jag dessutom äger en Kinect från tidigare, så hade jag mitt tema för examensarbetet klart.

## 1.2 Syfte och mål

Syftet med detta arbete är att undersöka hur användbara naturliga användargränssnitt är, jämfört med dagens fjärrkontroller. Jag kommer att reda ut hur processen för ansiktsgenkänning, röst- och rörelsestyrning fungerar i praktiken och hur välutvecklade dagens produkter med dessa egenskaper egentligen är samt vilka eventuella brister de har. Produkterna som jag kommer att koncentrera mig på är alltså Microsofts Kinect och Samsungs Smart TV.

Mina huvudsakliga forskningsfrågor är:

- Hur fungerar tekniken för ansiktsgenkänning, röst- och rörelsestyrning?
- Hur välutvecklat är det naturliga användargränssnittet i Kinect och Samsungs Smart TV, finns det några brister eller fungerar de väl?
- Hur användbara är naturliga användargränssnitt jämfört med dagens fjärrkontroller?

### 1.3 Avgränsning

Det finns många olika interaktionstekniker som gör att ett användargränssnitt kan räknas som naturligt, men jag kommer endast att behandla ansiktsigenkänning, röst- och rörelsestyrning. Mitt arbete avgränsas så att jag kommer att fokusera mig på Microsofts Kinect och Samsungs Smart TV, förutom då det gäller ansiktsigenkänning, röst- och rörelsestyrning på ett allmänt plan. Orsaken till att jag kommer att undersöka just de här två produkterna är att de är de enda välkända konsumentprodukterna i dag som innehåller alla tre funktioner: ansiktsigenkänning, röst- och rörelsestyrning. I detta examensarbete undersöker jag inte produkter vars rörelsestyrningsfunktioner kräver beröring eller en kontroll. Jag kommer endast att ta upp produkternas naturliga användargränssnitt, inte t.ex. hur Samsungs Smart TV annars fungerar som TV. Produkterna som jag undersöker är menade för vardagligt hemmabruk.

### 1.4 Metoder

Detta är till största delen frågan om en litteraturstudie. Det mesta materialet är hämtat från internet, eftersom det är där som den allra färskaste informationen oftast finns. Jag har fäst stor uppmärksamhet på källorna för att kunna basera utredningen på så pålitligt material som möjligt. Jag har även hittat några böcker som hör till ämnet, och gått igenom tekniska tidskrifter på jakt efter användbara produkttester.

Eftersom jag själv har tillgång till både Kinect och Samsungs Smart TV, så har jag haft möjlighet att utföra egna tester. I samband med testerna har jag gjort undantag och även förlitat mig på sådana källor som kanske inte anses vara av högsta klass när det gäller pålitlighet, men som är sakliga och innehåller intressanta observationer som jag har kunnat undersöka vidare i då jag utfört mina egna tester. Dessa tester har hjälpt mig att svara på frågorna gällande hur bra det naturliga användargränssnittet i produkterna fungerar och vilka eventuella brister de har, samt hjälpt mig besvara den huvudsakliga frågan om användbarheten.

## 1.5 Terminologi och förkortningar

**Ansiktsgenkänning:** En funktion som gör att en maskin kan känna igen en människa genom att studera ansiktsdragen.

**Algoritm:** Ett problem kan systematiskt lösas med hjälp av en algoritm som består av förutbestämda metoder för att räkna ut svaret.

**CMOS:** Står för Complementary Metal Oxide Semiconductor, och är ett chip som fungerar som en energisnål sensor som fångar upp bilder i en kamera.

**Fps:** Frames per second.

**IR:** Infraröd.

**Kinect:** Ett tillägg till Microsofts Xbox 360 som innehåller funktioner för ansiktsgenkänning, röst- och rörelsestyrning.

**Kognitiv belastning:** Betyder att något är ansträngande för hjärnan och minnet, d.v.s. det kräver mycket tankearbete.

**Maskininlärning:** Man gör en maskin intelligent genom att ge den träningsdata utgående från vilket den sedan lär sig att fatta beslut i nya liknande situationer.

**Natural User Interface, NUI:** Ett användargränssnitt där man använder sådana interaktionsmetoder som känns naturliga för en människa.

**OpenNI:** Mjukvara som finns på PS1080 som bl.a. känner igen rörelser.

**Peltier-element:** Ett element som har kontroll över temperaturen genom att höja eller sänka den.

**PS1080:** Ett chip som sköter om ljudinformationen och den visuella informationen i Kinect.

**Rörelsestyrning:** Innebär att man styr en maskin genom att använda rörelser.

**Röststyrning:** Innebär att man styr en maskin genom att tala till den.

**RGB-kamera:** En vanlig färgkamera där RGB står för röd, grön, blå.

**RGBD:** En färgbild och en djupbild som är sammansatt.

**Smart TV:** En TV som även fungerar som en dator då den har internetförbindelse.

**Smart Interaction:** Samsung Smart TV:ns funktioner för ansiktsigenkänning röst- och rörelsestyrning.

**SPSN:** Står för Samsung Product Support Network.

**Superkardioid:** En riktad mikrofon som tar upp den största delen av ljudet framifrån, men som även fångar ljud bakifrån.

**System-on-a-chip (SoC):** Innebär att de elektroniska funktionerna bildar ett ihopkopplat system på ett och samma chip.

**Träningsdata:** Data som används för att lära en maskin att fatta en viss sorts beslut och att på så sätt göra en maskin intelligent.

## 2 NATURLIG INTERAKTION

Vad innebär ett naturligt användargränssnitt? Ordet ”naturligt” kommer från det att användargränssnittet styrs med hjälp av både medfödda kunskaper och kunskaper som man har uppnått som följd av växelverkan med omgivningen. Ett sådant användargränssnitt styrs med andra ord på ett sätt som är naturligt för en människa. En medfödd kun-

skap är t.ex. det att man bland flera ljudkällor kan urskilja en talare och koncentrera sig just på denna eller det att man lägger märke till förändringar i sin omgivning. Tal är ett exempel på en inlärd kunskap. (Blake 2012 s. 9)

När det gäller naturliga användargränssnitt så strävar man till att återanvända färdigheter som redan existerar och som är enkla att använda. Inlärningsprocessen för sådana här kunskaper är enkel, de är en vidareutveckling av medfödda färdigheter och de har låg kognitiv belastning. Man kanske har lärt sig färdigheten redan efter att ha sett en exempel demonstration en gång och gjort ett par egna försök. Med att återanvända menas det här att man på ett enkelt sätt anpassar tidigare inlärd kunskaper för nya uppgifter i nya sammanhang. Om man följer dessa principer vid planeringen av det naturliga användargränssnittet så borde resultatet då vara ett användargränssnitt som användaren snabbt lär sig behärska och som känns naturligt att använda. Att använda en mus räknas inte som en naturlig interaktionsmetod eftersom den kognitiva belastningen är stor och inlärningsprocessen inte är den simplaste, och dessutom så är återanvändningsmöjligheterna få. (Blake 2012 s. 10-16)

Det finns många olika sorters funktioner som gör att ett användargränssnitt kan räknas som naturligt. Naturliga användargränssnitt kan byggas upp av olika typer av användargränssnittsteknologier, det är inte frågan om någon specifik nymodig teknik såsom t.ex. beröring eller rörelsestyrning. Det hela handlar om att interaktionen i huvudsak bygger på redan tidigare inlärd kunskaper, vilket gör den naturlig för användaren. (Blake 2012 s.) I detta examensarbete behandlas tre sådana funktioner som gör att ett användargränssnitt kan betraktas som naturligt: ansiktsgenkänning, röst- och rörelsestyrning.

## **2.1 Rörelsestyrning**

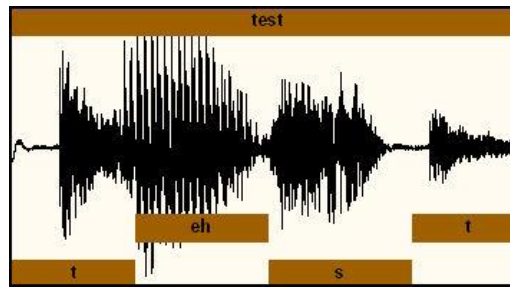
Rörelsestyrning innebär att man kommunicerar med en maskin genom handgester (såsom i Samsungs Smart TV) eller genom att utföra rörelser med någon annan del av kroppen (Panwar 2012 s. 1, Rai 2012). Vissa rörelsestyrningstekniker fungerar med rörelser från hela kroppen (exempelvis Microsofts Kinect) (Sung 2011 s. 93). Kontrollerfri videobaserad rörelsestyrning är mer naturlig och medför mer frihet då man inte behöver hålla i någon kontroll (såsom fallet är i Nintendo Wii och Playstation Move) eller

ha någon möjning fastkopplad på kroppen (Moni & Ali 2009 s. 433, Sung 2011). En eller flera kameror registrerar kroppsrörelserna, d.v.s. uppmärksammar ledens eller leder-  
nas ställning och position i rummet. Användningsutrymmet begränsas då till det område  
som kameran reagerar på. För att ett system överhuvudtaget skall kunna tolka kroppsrö-  
relser måste det först särskilja de leder som kroppsstyrningen fungerar med ur bilden.  
Detta ställer en hel del krav på systemet, inte minst med tanke på att rörelsestyrningen  
skall fungera i realtid och att rörelser skall registreras utan avbrott. Rörelseregistrerings-  
processen kan delas upp i två huvudsakliga delar: förbehandling som innebär bearbet-  
ning av bilden och sedan själva igenkänningen av en rörelse. (Moni & Ali 2009 s. 433)

Då en människa kommunicerar med en annan människa så är man sällan helt stilla.  
Speciellt handrörelser används ofta till hjälp då man vill uttrycka sig. Att kommunicera  
med rörelser är naturligt för en människa, och därför forskas det mycket i hur man kan  
tillämpa rörelser som kommunikationsmedel även mellan en människa och en maskin.  
(Moni & Ali 2009 s. 433)

## **2.2 Röststyrning**

Röststyrning, d.v.s. det att man får en maskin att utföra något genom att tala med den, är  
något som man har forskat i redan under flera tiotals år. Under den senaste tiden har  
forskningen gått en hel del framåt, men trots det är röststyrning fortfarande en utma-  
ning. Förutom termen röststyrning så använder man även begrepp såsom taligenkänning  
eller automatisk taligenkänning. (Fook 2012 s. 479) Det hela handlar huvudsakligen om  
att jämföra mönster och att finna sådana mönster som stämmer överens. Maskinen inne-  
håller en databas med förenklade representationer av varje ord. Det kan även finnas re-  
presentationer för var och ett av dessa ord i olika sammanhang. Systemet fångar upp tal  
och jämför det med orden i databasen och söker efter den närmaste motsvarigheten.  
Röststyrning är en utmaning av flera orsaker, exempelvis så har människor olika accen-  
ter och de talar med olika toner samt med olika ljudstyrkor. Själva mikrofonens egen-  
skaper måste även tas i beaktande. (Bladon 1994) Röststyrning fungerar ofta väl i en  
tyst omgivning, men vid buller brukar det uppstå problem (Fook 2012 s. 479).



Figur 1. Röststyrning handlar om att jämföra mönster (Rankin 2007).

Olika taligenkänningstekniker skiljer sig bl.a. beroende på om fortsatt tal, d.v.s. tal utan tystnader mellan orden kan tolkas, eller om taligenkänningsfunktionen endast klarar av ett enskilt ord åt gången. En annan skiljande faktor är om taligenkänningsfunktionen klarar av att känna igen tal från vem som helst eller om den måste skolas till att kunna tolka en användares röst. Storleken på maskinens ordförråd, karakteriserar även röstigenkänningstekniken och avgör mängden ord som maskinen kan reagera på. Ju fler ord ett system känner igen, desto naturligare känns det för användaren att använda systemet. (Bladon 1994)

### 2.2.1 Visuell taligenkänning

Då en människa lyssnar på vad en annan människa säger så följer man ofta även med munrörelserna, vilket gör tolkningen lättare. Speciellt i bullriga miljöer så underlättar det att följa med hur munnen rör sig, och exempelvis hörselskadade människor har stor nytta av detta. Visuella observationer kan på samma sätt underlätta taligenkänningen hos en maskin. För en maskin handlar det i stora drag om att upptäcka ansiktet och att lokalisera munnen och därefter att följa med läpprörelserna. Förutom att maskinen måste ta i beaktande hur munnen är formad så bör den också studera tänderna och tungan. (Fook 2012 s. 479-482)

## 2.3 Ansiktsigenkänning

Ansiktsigenkänning innebär att en maskin kan känna igen en människa utgående från ansiktsdragen. Det handlar i stora drag om att ett system fångar upp bilder av en person, lokaliserar ansiktet i bilden och söker därefter ut vissa beståndsdelar i ansiktet. Då dessa element har lokaliserats så placerar systemet utgående från dessa ansiktet i vad som an-

ses vara dess normalläge vad gäller vinkling, storlek och position. (Ban et al. 2008 s. 2408-2409) Efter det, då ansiktsbilden är behandlad så att den motsvarar en viss förutbestämd geometrisk modell, så kan bilden jämföras med en databas som innehåller tidigare sparade ansiktsbilder, i ett försök att hitta ett ansikte som så mycket som möjligt motsvarar den uppfångade bilden (Lone et al. 2011 s. 222). Jian Sun är en av forskarna som har jobbat med Kinects ansiktsgenkänningsfunktion. Han säger att det än så länge är omöjligt för en apparat att göra en säker identifiering, men att man kommer långt genom att låta ett system utföra gissningar baserade på genomtänkta hypoteser. (Gantenbein 2011)

Hur bra ett system är på att upptäcka och separera ett ansikte från en uppfångad bild har direkt inverkan på hur bra ansiktsgenkänningsfunktionen fungerar. En apparat bör därför så noggrant som möjligt kunna registrera gränserna för ett ansikte och separera ansiktet från resten av bilden. Tillvägagångssätten för en maskin att upptäcka ett ansikte kan indelas i två huvudkategorier: metoder där egenskaper tas i beaktande och metoder där ansiktets former tas i beaktande. Registrering av hudfärg är ett exempel som hör till den förstnämnda metoden för att upptäcka ansikten. (Ban et al. 2008 s. 2408)

### **2.3.1 Ansiktsgenkänning på ett längre avstånd**

Då en användare befinner sig på ett längre avstånd från kameran så kommer ansiktsbilden som registreras att bli liten. Då bilden sedan normaliseras till en större storlek så kommer den att bli en aning suddig vilket leder till att det blir svårare för systemet att se likheterna då ansikten jämförs. Det här problemet kan lösas genom att man lägger på ett filter på de sparade bilderna i databasen, som då även gör dem suddiga och på så sätt blir det lättare för systemet att para ihop den uppfångade ansiktsbilden med en bild i databasen. (Ban et al. 2008 s. 2409-2410)



*Figur 2. Hur avstånd kan representeras genom användning av filter (Ban et al. 2008 s. 2410).*



### **3 KINECT**

Microsofts Kinect är ett tillägg till Microsofts spelkonsol Xbox 360 som erbjuder en kontrollerfri videospelupplevelse (Cruz et al. 2012 s. 36). Kinect är en sensor med vilken man kan styra menyerna och själva spelen genom att röra på olika delar av kroppen. Den klarar även av röststyrning och ansiktsigenkänning. Kinect är ett resultat av ett samarbete mellan Microsoft och PrimeSense. (Zhang 2012 s. 4-5) Sensorn, som från början presenterades som "Project Natal", kom till försäljning i november år 2010 (Cruz et al. 2012 s. 36-37). En kort tid efter att den lanserades så utnämndes den till den snabbast säljande konsument elektronikpylen någonsin, och den finns därför med i Guinness rekordbok. Även om Kinects användningsområde i dagens läge är mycket brett, så kommer jag i detta examensarbete endast att koncentrera mig på dess ursprungliga användningsändamål, som tillägg till Xbox 360 för kontrollerfri styrning av videospel i hemmet. (Zhang 2012 s. 4-5)

#### **3.1 Komponenterna**

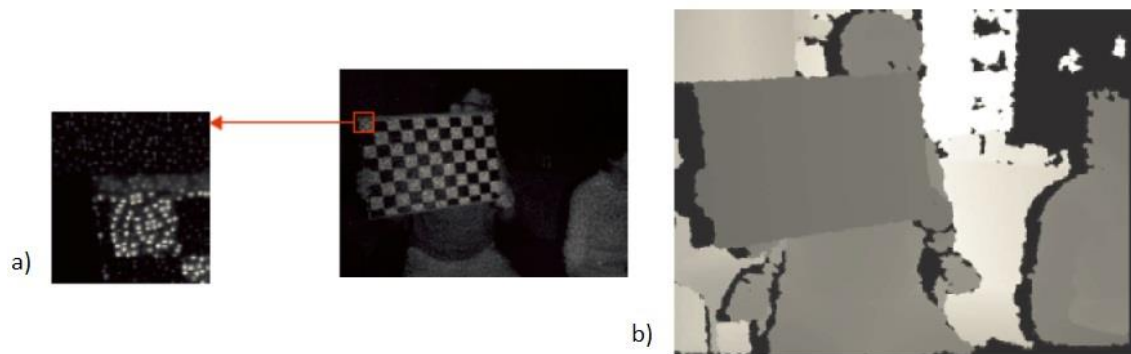
Kinect är till utseendet en avlång, vågrät apparat som är placerad på en fot (Zhang 2012 s. 5). Stativet är motorstyrt så att huvudet kan vändas uppåt eller neråt tills det är i en lämplig position i förhållande till användarna framför apparaten. En accelerometer håller reda på hur huvudet är riktat, genom att ta gravitationen i beaktande och räknar ut vad som är neråt. (Kramer et al. 2012 s. 23) Kinect innehåller en vanlig RGB-kamera. Två andra huvudkomponenter är en IR-projektor (d.v.s. en infraröd laserprojektor) och en IR-kamera, som tillsammans med mjukvaran i Kinect bildar ett system som gör att omgivningen kan uppfattas i 3D. Färgkameran och den IR-kameran är båda s.k. CMOS-kameror. (Forsythe & Green 2012 s. 1-2) Nyckelkomponenten i Kinect är PrimeSenses "system-on-a-chip" eller systemchip: PS1080, som hanterar både den visuella och den auditiva informationen. Mikrofonerna i Kinect är fyra till antalet. (How Microsoft's PrimeSense-based Kinect really works 2011)



Figur 3. Hur Kinect ser ut utanpå (Kramer et al. 2012 s. 2).

### 3.2 Hur djupmätningen fungerar

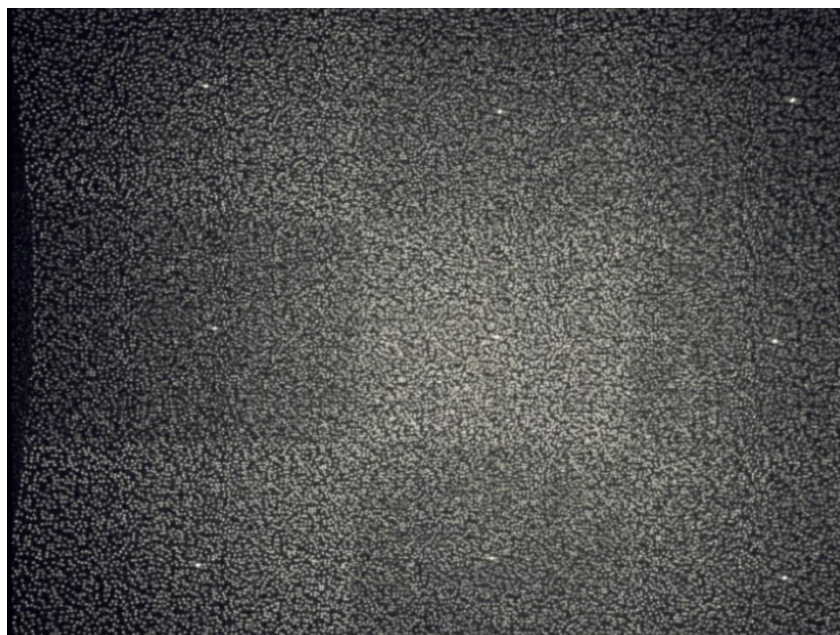
Den infraröda kamerans och den infraröda laserprojektorns roll i djupmätningen handlar om strukturerat ljus. Denna princip bygger på att laserstrålen, som IR-projektorn strålar ut i rummet på 830 nm, bildar ett mönster av ett åtskilligt antal av små prickar som sedan uppfattas av IR-kameran. (Kramet et al. 2012 s. 11-12) Kinect har ett referensmönster i sitt minne, d.v.s. ett standardmönster att utgå från (Khoshelham 2011 s. 133). På det sättet så vet Kinect den ”normala” storleken och placeringen av laserprickarna, samt då även deras avstånd till kameran (How Microsoft's PrimeSense-based Kinect really works 2011 s. 28). Då Kinects laser sedan träffar objekt i rummet så bildas det avvikelser i mönstret som uppstår jämfört med standardmönstret. Utgående från prickarna kan apparaten sedan avgöra objektens avstånd och placering i rummet, och därmed bygga upp en 3D modell av omgivningen. Det är dock inte IR-sensorn och IR-kameran som uppfattar omgivningen som tredimensionell. Deras uppgift är endast att skapa och fånga upp prick-mönstret. IR-kamerans CMOS-sensorer skickar sedan informationen vidare till PS1080-chippet. Med hjälp av denna information och ytterligare information från RGB-kamerans CMOS-sensorer, så kan chipet med algoritmer bygga upp en tredimensionell bild av rummet. (Forsythe & Green 2012 s. 2-3) Chipet tolkar IR-prickarna med en hastighet på 30 fps (How Microsoft's PrimeSense-based Kinect really works 2011 s. 28).



Figur 4. a) IR-prickar kan ses i den vänstra bilden som är en närbild av området i den röda rutan till höger  
b) djupbild (Zhang 2012 s. 5).

### 3.2.1 IR-laserprojektorns egenskaper

För att djupmätning med strukturerat ljus skall fungera korrekt så måste laserns våglängd alltid vara den samma. Detta försöker man åstadkomma med ett peltier-element, som strävar till att hålla laserdiodens (och på samma gång den utstrålade våglängdens) temperatur oförändrad, genom att höja och sänka temperaturen på laserdioden. IR-laserns prickmönster som täcker rummet består av nio medelstarka laserpunkter omringade av mindre svagare punkter. Dessa nio punkter är ett resultat av den ofullständiga ljusfiltreringen som används till att uppnå prickmönstret. Då ljus filtreras på det här sättet för att få ett prickmönster, så är det vanliga resultatet att det bildas en extra stark ljuspunkt i mitten. PrimeSense har utvecklat och patenterat tekniken för att ersätta en enda stark punkt med nio svagare punkter. Eftersom ljuset från lasern uppdelas på detta sätt så har man kunnat förse Kinect med en starkare laser. Filtret som gör att lasern från IR-projektorn bildar ett prickmönster, begränsar dock laserns våglängd till 830 nm av den orsaken att lasern skall vara trygg för ögonen. Utan filtret ligger effekten vid 70 mW och lasern skulle då vara farlig för ögonen. (Kramer et al. 2012 s. 11-13)



*Figur 5. IR-laserns prickmönster där de nio starkare punkterna går att urskilja (Kramer et al. 2012 s. 12).*

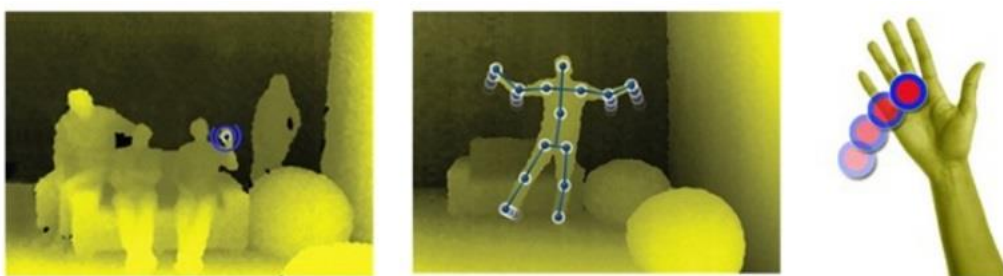
Styrkan på lasern är den faktor som avgör det maximala avståndet från Kinect som dess sensorer reagerar på (Kramer 2012 s. 12). Området som Kinect uppfattar börjar vid 0,8 meter från apparaten och det längsta avståndet den avläser ligger ungefär 3,5 meter från kameran. Tanken är att ett område av den här storleken skall räcka till för att utföra de rörelser som Kinects videospel kräver. (How Microsoft's PrimeSense-based Kinect really works 2011 s. 28)

### **3.2.2 Störningar från omgivande ljus**

Det omgivande ljuset i rummet är ett problem vid djupmätningen, då det kan blända sensorn eller förorsaka en felaktig djupmätning. Den IR-kameran har ett filter för genomsläppning av infrarött ljus som går vid 830 nm våglängd för att minska på störningar från olika ljuskällor i omgivningen. Tack vare filtret så reagerar inte kameran på IR-ljus som ligger vid andra våglängder, såsom ljus från en fjärrkontroll. Filtret eliminerar dock inte alla ljusproblem såsom solljus, eftersom solljuset har så mycket inverkan på 830 nm våglängd att det kan störa djupmätningen och blända sensorn. (Kramer et al. 2012 s. 12)

### 3.2.3 RGBD

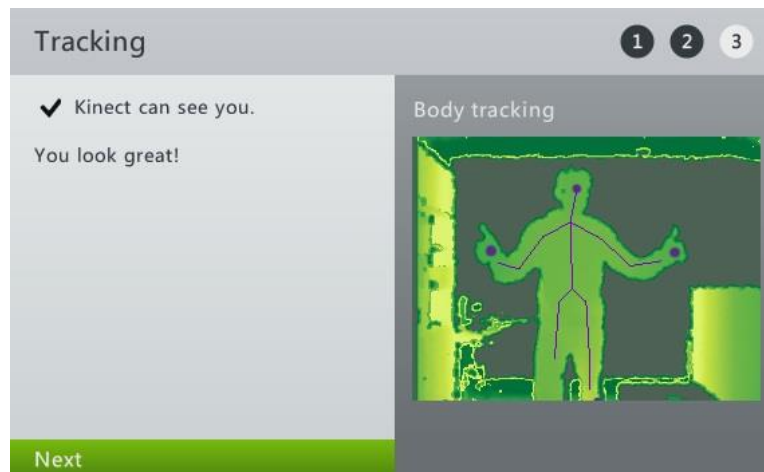
Kinects PS1080-chip sammanställer färgbilden från RGB-kameran med djupbilden som fås från IR-kameran till en RGBD-bild, där D står för djup. Detta lyckas lätt eftersom färgkamerans och IR-kamerans CMOS-sensorer är placerade sida vid sida. Mjukvaran OpenNI som finns på chippet får sedan tillgång till denna sammanfogade RGBD-information. Programmet kan bl.a. känna igen rörelser. (Forsythe & Green 2012 s. 3) Tack vare RGBD-uppgifterna så kan Kinect tolka t.ex. en fot-ben-kropp relation (How Microsoft's PrimeSense-based Kinect really works 2011 s. 30).



Figur 6. OpenNI kan bl.a. känna igen en hand, bearbeta skelett och känna igen rörelser (Cruz et al. 2012 s. 42).

## 3.3 Skeletal-tracking och hur rörelsestyrningen fungerar

Kinect kan reagera på rörelser från hela kroppen, oberoende av kroppsbyggnad. Rörelsestyrningen i Kinect bygger på de 3D-bilder som apparaten framställer. I dem kan systemet särskilja en persons siluett från bakgrunden och sedan används en algoritm för att indela siluetten i dess olika kroppsdelar. Därefter drar systemet slutsatser om var denna persons leder ungefär ligger. (Knies 2011b) Detta gör Kinect med en hastighet på 200 fps (Zhang 2012 s. 7). Lederna (som representeras i 3D-format) kopplas ihop till ett fullständigt skelett och sedan kan den virtuella kroppen som Kinect har producerat styras med den egna riktiga kroppen. Denna process var Kinect spårar en persons leder för att bygga upp ett hanterbart skelett kallas för "skeletal-tracking". (Knies 2011b)



Figur 7. Skeletal-tracking i Kinect. Djupmappens gråa nyanser som representerar avstånd färgas med gröna toner, eftersom grönt förknippas med Xbox. (Michel 2011)

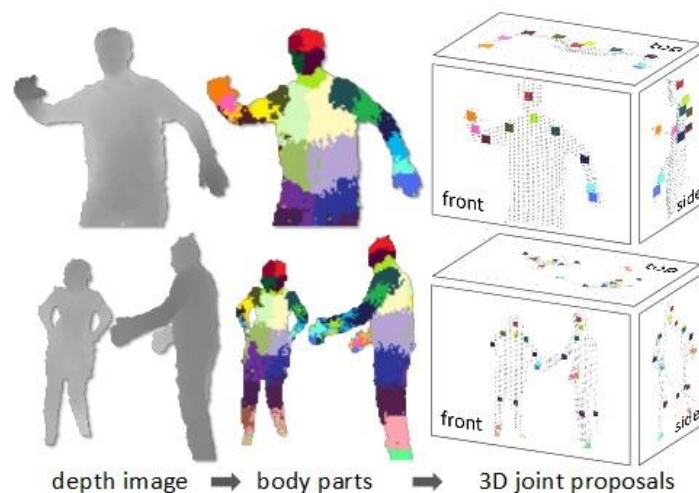
### 3.3.1 Kinect tolkar varje led skilt för sig

Idén som man utgick från under utvecklingsskedet av Kinects skeletal-tracking var att då en människa klarar av att rita in lederna på en bild av en person, så är det även möjligt att få en maskin att klara av uppgiften. Kinect behandlar varje inkommande bild skilt för sig, och bestämmer vid var och en av dem var en persons leder ungefär ligger. Systemet behandlar de olika lederna skilt för sig. Om kroppen direkt skulle behandlas som en helhet, så skulle varje möjlig kroppsställning och kroppsbyggnad vara tvungen att finnas representerad i en databas. De inkommande bilderna skulle sedan jämföras med varje bild i databasen. Då man i Kinects fall i stället jämför en led i taget så minskar antalet bilder i en sådan databas. Om man som exempel tänker sig att en hand kan vara placerad på tre olika sätt, att en handled också kan vara placerad på tre olika sätt, och att en arm kan inta två olika ställningar, så skulle de tillsammans kunna inta 18 olika ställningar ( $3 \cdot 3 \cdot 2$ ). Om de däremot behandlas skilt för sig så kan de inta 8 olika ställningar ( $3 + 3 + 2$ ), vilket minskar på datamängden. (Knies 2011b)

### 3.3.2 Objektigenkänning och per-pixel klassificering

Hur kan då Kinect räkna ut vilka områden den skall behandla som skilda leder? För att lösa detta används s.k. objektigenkänning. Med hjälp av objektigenkänningsfunktionen så kan Kinect räkna ut vad som t.ex. är en hand och vad som är ett huvud i en bild och sedan färglägga dessa kroppsdelar med olika färger för att skilja dem åt. Systemet är

programmerat så att de olika färgerna ger vägledning om var lederna ligger. (Knies 2011b) Då Kinect får en gråskalig bild från IR-kameran så behandlar den varje pixel i bilden skilt för sig och drar slutsatser om vilka pixlar som hör till vilken kroppsdel. Denna process kallas för per-pixel klassifikation. (Zhang 2012 s. 7) De olika nyanserna av grått i bilden från IR-kameran beskriver djupet. Om man då exempelvis färgar alla pixlar som hör till en Kinect-användares högra fot orange, så att de bildar en helhet, så kan man med denna information tillsammans med djupinformationen framställa högra fotens XYZ-koordinater. (Knies 2011b)



Figur 8. En djupbilds pixlar färgläggs i olika färger beroende på vilken kroppsdel varje enskild pixel tillhör. Därefter kan Kinect räkna ut XYZ-koordinaterna för kroppens leder. (Shotton et al. 2011 s. 1297)

### 3.3.3 Maskininlärning och träningsdata

Kinect koncentrerar sig på enskilda kroppsställningar och inte på direkta rörelser (Shotton et al. 2011 s.1299). Apparaten kan spåra skelett tack vare maskininlärning. Till det behövs träningsdata, d.v.s. data som kan användas till att skola maskinen så att den lär sig att räkna fram sig till rätt svar. Det skulle dock vara alltför tids- och kostnadskrävande att samla ihop träningsdata från riktiga bilder där man namnger människors kroppsdelar manuellt. Man skulle då inte heller få ihop den mängd data som man behöver. Det går däremot effektivt att skapa konstgjorda djupbilder med datorgrafik vilket man har gjort i Kinects fall. I samband med skapandet av dessa djupbilder används en mjukvara som kan tillämpa färdiga rörelsedatauppsättningar på olika individer i varierande kroppsställningar. (Knies 2011b)



Att datamässigt få med alla poseringar som en människa kan inta skulle vara en svårlöst uppgift. Kinect behöver dock inte känna till var eviga en kroppsställning, eftersom systemet till en del klarar av att dra egna slutsatser vid okända ställningar. Olikheter i fråga om kroppsbyggnad, var en person befinner sig i rummet, klädsel m.m. är sådant som systemet kan ta i beaktande utan att man behöver producera sådana skillnader med rörelseprogram. Skillnader som finns i närliggande ramar av rörelsedata är för det mesta så betydelselösa att man har kunnat lämna bort en hel del data, med det slutresultatet att två olika ställningar skiljer sig till åtminstone 5 cm. (Shotton et al. 2011 s. 1299) ”Exemplar” är namnet på den effektiva algoritmen som fungerar så snabbt att den går att tillämpa på varje inkommande bild från IR-kameran. En annan algoritm använder sedan informationen som fås från Exemplar-algoritmen för att skapa ett komplett skelett av de olika lederna, så att de därefter kan behandlas som en fungerande men dock tillfällig helhet. (Knies 2011b)

### **3.3.4 Varför basera rörelsestyrningen på djupmätning?**

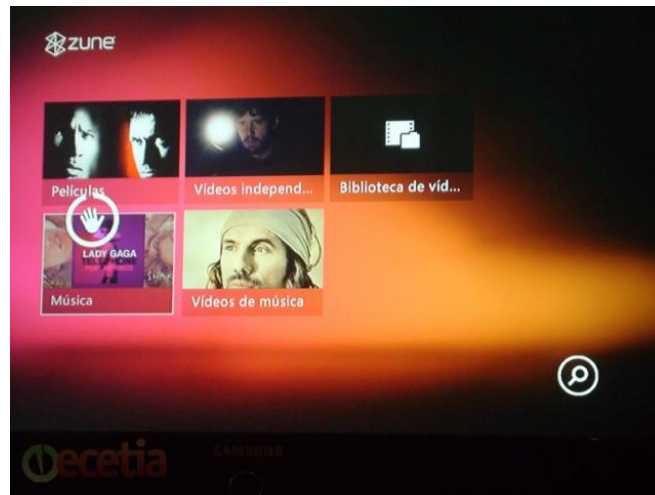
Förutom att man med en djupkamera bekymmerfritt får ihop ordentligt med träningsdata p.g.a. att det med en sådan kamera effektivt går att bygga upp konstgjorda kroppsstrukturer som ser äkta ut, så har kameran även andra fördelar. Den stöter exempelvis inte på problem på grund av vissa färger eller texturer i omgivningen och den klarar av sin uppgift även i mindre upplysta rum. Den kan även lätt urskilja en person och ”klippa ut” personen ur bakgrunden, samt bearbeta kroppsställningar även om de består av oklarheter (såsom att en kroppsdel ligger en aning ytterom kamerans synfält). (Shotton et al. 2011 s. 1298-1299)

### **3.3.5 Hur man använder rörelsestyrningen i Kinect**

Rörelsestyrningen i Kinects meny och i de flesta av apparatens spel fungerar så att man sträcker fram handen, varefter en ikon som föreställer en hand visas på TV-skärmen. Denna ikon rör sig enligt de handrörelser man gör, så det gäller då att med handen styra ikonen ovanpå det objekt som man vill välja. Då ikonen befinner sig över ett alternativ så omringas handikonen av en förloppsindikator, som visar hur länge man skall hålla



handen på samma ställe för att välja i frågavarande alternativ. Dance Central, är ett exempel på ett spel där man väljer sina alternativ i menyn genom svep-rörelser, d.v.s. genom att dra handen från vänster till höger eller tvärtom. (Rivington 2010) Då man börjar spela ett spel så blir rörelsestyrningsfunktionen mångsidigare, eftersom också andra kroppsdelar förutom händerna, om inte t.o.m. hela kroppen, skall användas till att styra spelet (Zhang 2012 s. 4).



*Figur 9. En förloppsindikator visar hur länge man skall hålla handikonen vid ett objekt för att välja det (Mejia 2010).*

### 3.4 Hur röststyrningen fungerar

Kinect innehåller 4 stycken superkardiodida mikrofoner (Tashev 2012). Dessa uppfattar ljud inom ett brett område framför apparaten. Kinect kan räkna ut från vilken riktning ljudet kommer och den kan även urskilja olika ljudkällor då det finns en blandning av ljud i omgivningen. Kinect bör kunna urskilja ljud noggrant på det här sättet för att inte störas av ljudet från TV:n eller av annat ljud i dess närhet, och för att kunna uppfatta kommandon då flera personer talar på en gång. För att möjliggöra detta så genomgår ljudet specialiserade luddämpningsprocesser. På samma sätt så som kamerorna sänder information som elektriska signaler till PrimeSenses PS1080-chip, så sänds i sin tur även informationen som mikrofonerna fångar upp till chippet i form av elektriska signaler. Mjukvaran i PS1080-chippet behandlar informationen från både kamerorna och mikrofonerna och synkroniserar även all denna information. (Forsythe & Green 2012 s. 2) Då informationen är synkroniserad så kan Kinect jämföra det den ser med det den hör

och på så sätt räkna ut om det är användaren som talar eller om ljudet kommer någon annanstans ifrån (Naone 2011 s. 83).

### **3.4.1 Ljudbehandling**

En väsentlig sak som måste tas i beaktande vid röststyrning är ljud från högtalare. Kinect med dess mikrofoner ligger närmare TV-högtalarna än till Kinect-användarna. Ett talat kommando blir då svårt att uppfatta över ljudet som kommer från högtalarna. Samma problem finns vid användningen av en högtalarfunktion under ett telefonsamtal. Som lösning på denna problematik används i vanliga fall en algoritm som går under namnet "acoustic echo cancellation". Samma algoritm fungerar inte i Kinects fall eftersom ljudet från högtalarna ofta är högre jämfört med ljudet från en person som talar och eftersom användarna uttalar kommandon från ett längre avstånd till mikrofonerna. Dessutom så används det endast en överföringskanal till ljudåtergivningen från en vanlig högtalartelefon, medan en TV-apparat har stereo ljud eller eventuellt surround ljud. Kinect klarar av att dämpa ekon av en högre ljudstyrka och behandla efterklang, samt använder algoritmen "stereo acoustic echo cancellation". Dessa faktorer bidrar till att systemet kan uppfatta talade kommandon från olika riktningar på ett avstånd upp till fyra meter i ett normalt bullrigt vardagsrum. (Knies 2011a)

## **3.5 Kinect Identity**

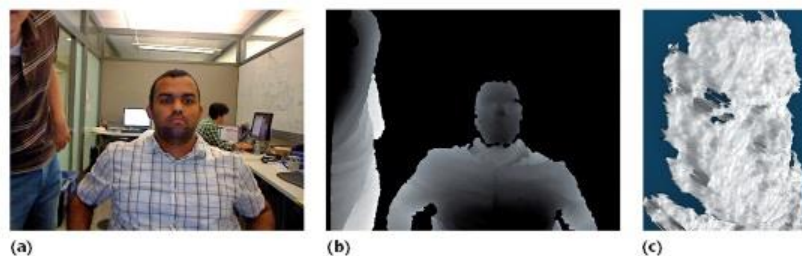
Till Kinects naturliga användargränssnitt hör även en ansiktsgenkänningsfunktion. Kinect tar ansiktsgenkänningsfunktionen ett steg längre och studerar förutom ansiktsdragen dessutom färgen på kläderna och längden på en person, i hopp om att få ett pålitligare resultat. Ansiktsgenkänningen är dock den viktigaste delen av identifieringsprocessen. (Leyvand et al. 2011 s. 94) Denna process med alla dess delmoment för att känna igen användare kallas för Kinect Identity (Gantenbein 2011).

Personidentifieringen i Kinect kan uppdelas i "biometrisk inloggning" och "session". Den biometriska inloggningen handlar om identifikation på lång sikt, där en persons egenskaper läggs i minnet på Kinect så att apparaten i fortsättningen kan logga in denna användare, då denna befinner sig inom kamerans synfält. Session-delen är däremot

kortvarig och finns till för att hålla reda på olika spelare under en spelsession. (Leyvand et al. 2011 s. 94)

### 3.5.1 Ansiktsigenkänning

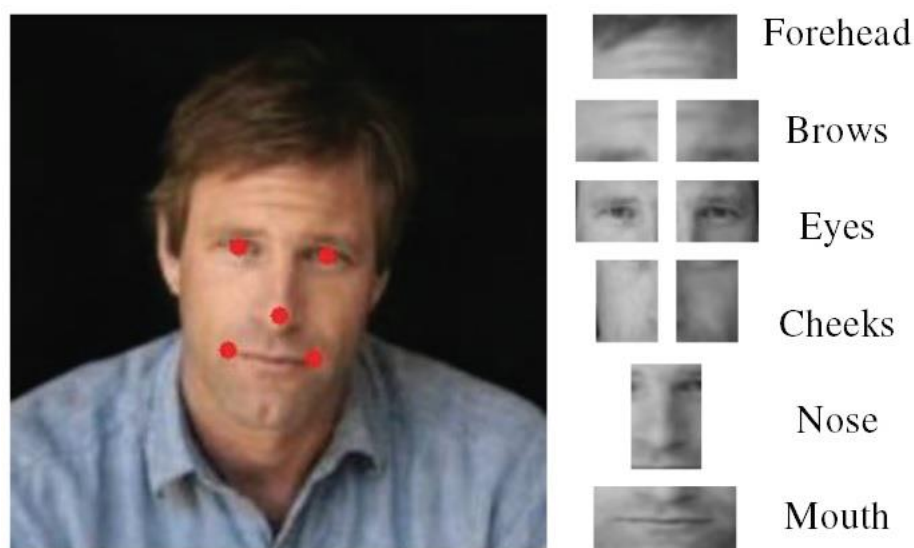
Ansiktsigenkänningsfunktionen i Kinect är uppbyggd så att den är flexibel och skall fungera trots växlande omständigheter. Förutom att utseendet ändras då man gör olika miner, så ser ett ansikte annorlunda ut även i t.ex. olika ljus och vid olika avstånd från kameran. (Gantenbein 2011) Kinects ansiktsigenkänningsprocess består av tre delar. Ett delmoment är att registrera var ansiktet finns och notera dess storlek. (Leyvand et al. 2011 s. 94) De ansiktsbilder som systemet fångar upp av användaren förminskas eller förstoras, beroende på användarens avstånd till kameran. På så sätt uppnås den rätta storleken för att Kinect skall kunna behandla bilderna vidare. (Gantenbein 2011) Ett annat delmoment i ansiktsigenkänningsprocessen innebär att ansiktet måste ställas in så att det är i dess ”rätta” position, vilket betyder att huvudet är rakt och att ansiktet är riktat rakt mot sensorn. Det tredje delmomentet handlar sedan om den detaljerade identifikationen, där ansiktets mikrostruktur tas i beaktande. (Leyvand et al. 2011 s. 94)



Figur 10. Hur Kinect ser ett ansikte. a) vanlig RGB-bild b) djupbild c) ansiktets struktur (Zhang 2012 s. 8).

Kinect identifierar ansikten genom att fokusera sig på nio huvudpunkter i en persons anletsdrag. (Gantenbein 2011) Dessa är munnen, näsan, högra och vänstra ögat, högra och vänstra ögonbrynet, pannan, samt högra och vänstra kinden (Cao et al. 2010 s. 2712). Vissa av dessa nio element är tydligare än andra beroende på ljusförhållanden och hur huvudet är riktat, så Kinect koncentrerar sig då på några av de tydligaste delarna och baserar sedan sina gissningar på dem (Gantenbein 2011). I stället för att koncentrera sig på att jämföra ansiktet som en helhet, såsom många andra ansiktsigenkänningsfunktioner gör, så kan Kinect alltså ta olika delar av ansiktet i beaktande skilt för sig.

Chanserna för att åstadkomma en korrekt identifiering ökar då ansiktsdragen delas upp i olika bitar på det här sättet. (Cao et al. s. 2711-2712) Bilderna genomgår en filtreringsprocess som avlägsnar variationer förorsakade av ljus, varefter de behandlas vidare av mjukvaran. Med hjälp av en algoritm kan systemet sedan avgöra huvudets ställning, d.v.s. hur en användares ansikte är riktat. Efter det då alla delmoment är avklarade, så kan Kinect börja jämföra bilderna av de synliga huvudelementen i ansiktet med dem som den har i minnet, och därefter få fram den mest sannolika identiteten. Det sägs att Kinects ansiktsidentifiering i de flesta fallen lyckas till ungefär 85 procent. (Gantenbein 2011)



*Figur 11. Kinects ansiktsigenkänningsfunktion tar olika delar av ansiktet i beaktande skilt för sig (Cao et al. 2010 s. 2712).*

### **3.5.2 Personidentifiering**

Under en spelsession då en spelare ställer sig framför TV:n och Kinect, så befaller det i fråga varande spelet Kinect att ta bilder av denna person. Bilderna kommer sedan att analyseras och ansiktets struktur, färgen på kläderna och personens längd kommer att sparas i minnet så att Kinect lär sig att känna igen spelaren. Bilderna kommer också genast att jämföras med de bilder som redan finns i minnet så att Kinect sedan kan meddela om den känner igen personen från tidigare. (Leyvand et al. 2011 s. 94)

Hur fungerar identifieringen då i praktiken? Kinect avgör om den tror sig känna igen en person genom att använda en "sanningsvärdetabell". Tabellen består av alla spelare som finns i minnet och då alltså information om dessa spelares ansiktsdrag, kroppslängd och klädsel. Den oidentifierade personens ansiktsdrag jämförs med alla sparade identiteters ansiktsdrag. Om systemet anser att det funnit ett likadant ansikte bland spelarna som det har i minnet, så blir värdet vid "ansikte" positivt för denna identitet i sanningsvärdetabellen. Sådana identiteter som systemet anser att inte passar ihop med de uppmätta egenskaperna får ett negativt värde vid "ansikte" i tabellen, och i sådana fall då systemet inte kan avgöra om egenskapen kan kopplas ihop med en identitet så sätter den värdet till okänt. Efter att ansiktsdragen har jämförts med varje känd identitet så kommer klädfärgen att jämföras med klädfärgen för alla sparade identiteter och till sist jämförs längden på motsvarande sätt. Om en identitet i sanningsvärdetabellen inte har ett enda negativt värde och åtminstone ett positivt värde, så tolkar Kinect det som så att den känner igen spelaren. Om de uppmätta ansiktsdragen tydligt motsvarar ansiktsdragen hos en identitet som de jämförs med, så kan systemet göra undantag och anse att det funnit en träff trots negativa värden vid klädfärg och längd för denna identitet. Detta gäller dock endast för ansiktsdragen, eftersom det är den enda av de tre egenskaperna som går att beakta självständigt. Om fler än en identitet får ett positivt värde vid en egenskap så räknas värdena som om de skulle vara okända, eftersom resultatet inte anses vara pålitligt då en spelare bör kopplas ihop med endast en av de sparade identiteterna. (Leyvand et al. 2011 s. 94-95)

*Tabell 1. Sanningsvärdetabell. Kinect kommer här att anta att identitet nr 1 är den rätta eftersom den inte har negativa värden och minst ett positivt värde. (Leyvand et al. 2011 s. 95)*

Table 1. A truth table for identifying a new player in a game.				
Characteristic	Enrolled ID #1	Enrolled ID #2	Enrolled ID #3	Enrolled ID #4
Face	Positive	Negative	Positive	Unknown
Clothing color	Unknown	Unknown	Negative	Negative
Height	Positive	Positive	Unknown	Unknown

Med längdegenskapen kan man enkelt eliminera möjliga identiteter, men eftersom det kan finnas flera spelare i minnet som har samma mått så kan inte Kinect bara ta denna egenskap i beaktande vid personidentifieringsprocessen. Då fler egenskaper tas i beaktande så ökar alltså chanserna att åstadkomma en korrekt identifiering. Man kanske frå-

gar sig hur systemet bär sig åt för att känna igen en spelare vars utseende har ändrats sedan senaste spelsession? Det kan vara frågan om förändringar gällande t.ex. frisyr, skäggväxt eller användning av glasögon. Under en och samma spelsession så är sådana här förändringar inte så vanliga, men man ändrar däremot miner, kroppsställning och annat då man spelar. Identifieringsfunktionen bör därför vara flexibel. För att identifieringen skall fungera trots att någon förändring inträffar så tar Kinect detaljerna i rummet i beaktande. Detta gäller vid session-identifiering. Då sparas information om belysning och positioner i rummet. Då märkbara förändringar inträffar så sparas ny information som underlättar identifieringen. Vid den biometrisk inloggningen så tas endast ansiktsdragen i beaktande. Kinect sparar ett flertal olika bilder av en person i minnet för att ha varierande material att jämföra med och på det sättet göra identifieringsuppgiften lättare. (Leyvand et al. 2011 s. 95)

## **4 SMART TV**

Det naturliga användargränssnittet i den åttonde serien av Samsungs Smart TV kallas för "Smart Interaction" (Mossberg 2012). Med det kan man hantera menyer och vissa funktioner. Ansiktsgenkänning och rörelsestyrning fungerar bäst då man står på ett ungefärligt avstånd mellan 1,5 m till 4 m, och röststyrningens rekommenderade avstånd ligger mellan ungefär 1 m till 4 m. (Samsungs e-Manual 2012) Uppe på TV-apparaten finns det en inbyggd kamera och en mikrofon (Katzmaier 2012). Under ett test av Samsungs Smart TV som utfördes av Samsung Product Support Network, SPSN, så poängtedes det att det naturliga användargränssnittet gör användningen roligare och att man antagligen nog finner intresse för att använda de funktioner som räknas dit. I ett skede då man förevisar rörelsestyrningen så används däremot fjärrkontrollen emellan med den förklaringen att man gör det för att det går snabbare på det sättet. (Jordan 2012a) Hur användbart är då egentligen det naturliga användargränssnittet och hur mycket använder man det sist och slutligen, då man redan i en kort instruktions- och testvideo som denna övergår till att använda fjärrkontrollen för att det går enklare, snabbare och kanske mer problemfritt?



Figur 12. Samsungs Smart TV ES8005 (Benzon 2012).

## 4.1 Rörelsestyrning

Rörelsestyrningen i Samsungs Smart TV fungerar med handrörelser. Eftersom kameran inte reagerar på rörelser från hela kroppen såsom Kinect-sensorn gör, så är det ingen skillnad om man står eller sitter då man vill använda funktionen. (Jordan 2012a) Rörelsestyrningen kan bl.a. användas då man vill justera volymen, surfa på nätet, ändra på inställningar i menyn eller spela spel (Jordan 2012a, Mossberg 2012). Innan man börjar använda denna funktion så bör man köra ett test som kontrollerar att omgivningen är lämplig för rörelsestyrning. Det första som granskas är att det är tillräckligt ljust i omgivningen så att TV:n kan uppfatta de rörelser som görs, och om det inte är det så får man ett meddelande om att man bör öka på belysningen. (Jordan 2012a) Både vid användning av rörelsekontroll och vid ansiktigenkänning så rekommenderas det att belysningen ligger mellan 50 till 500 lux. Testet undersöker även om kameran är riktad mot användaren. (Samsungs e-Manual 2012) Om det framgår att kameravinkeln är fel, så kan man manuellt rikta kameran uppåt eller neråt genom att snurra på ett hjul som finns på baksidan av kameran (Jordan 2012a).

### 4.1.1 Hur används rörelsestyrningsfunktionen?

Rörelsestyrningsfunktionen måste aktiveras varje gång man vill använda den. Den kan inte vara aktiverad konstant eftersom vanliga rörelser man gör framför TV:n då skulle tolkas som kommandon. (Katzmaier 2012) För att fånga TV:ns uppmärksamhet och påbörja rörelsestyrningen så håller man handen och fingrarna raka och riktar handflatan

rakt mot TV:n. Sedan så för man handen några gånger från sida till sida som om man skulle vinka och efter det får handen kontroll över TV:n. Det rekommenderas att man utför långsamma och jämna rörelser för att systemet skall hänga med. För att exempelvis ”klicka” på något så stänger man den öppna handen en gång och öppnar den igen och för att gå tillbaka, t.ex. då man bläddrar bland meny- eller internetsidor, så för man den öppna handen motsols i en rund cirkel. (Jordan 2012a) Om den hand som har kontroll över TV:n sänks eller på annat sätt hamnar utanför området som kameran uppfattar, eller om någon börjar använda fjärrkontrollen eller eventuellt ett tangentbord, så avbryts rörelsestyrningsfunktionen (Samsungs e-Manual 2012). Om det händer att flera personer på samma gång försöker styra TV:n med rörelser, så kommer TV:n att följa den personen som först lyckas fånga dess uppmärksamhet (Jordan 2012a). I TV:ns användarmanual varnas det för fysisk utmattning vid längre bruk (Samsungs e-Manual 2012). Att man blir trött i armen är ingen sällsynt företeelse då man använder användargränssnitt som styrs med rörelser och det brukar kallas att man får ”gorilla arm” (Ionescu et al. 2011 s. 160).

#### **4.1.2 Ljuset och bakgrundens inverkan på rörelsestyrningen**

Rörelsestyrningen fungerar bäst om TV:n är placerad i ett rum som är ordentligt upplyst och då en lämplig mängd ljus träffar framsidan av användaren. Om det är väldigt ljust bakom användaren så kan kameran ha svårt att uppfatta personen och direkt solljus kan också förhindra kameran från att se användaren. Reflektiva bakgrunder bör även undvikas och det är bra om bakgrundsfärgen är av tillräcklig kontrast till användaren eftersom kameran letar efter rörelser. (Jordan 2012a)

### **4.2 Röststyrning**

Samsungs Smart TV innehåller både en inbyggd mikrofon på själva TV:n och en annan på dess Smart Touch-fjärrkontroll vilka kan användas till röststyrning (Samsungs e-Manual 2012). TV:ns röststyrningsfunktion kan användas till att bl.a. ändra volym och till att navigera till olika program såsom Smart Hub eller Webbläsaren (Jordan 2012c). Innan man börjar så lönar det sig att köra ett test som undersöker omgivningen och dess egenskaper. Apparaten kommer då först att kontrollera hur mycket ljud det förekommer



i rummet. Det rekommenderas att ljudet i omgivningen är under 40 dB för att röststyrningen skall fungera bra. Efter bullertestet kommer ett mikrofon- och högtalartest att utföras. (Samsungs e-Manual 2012) Ljudet i omgivningen kommer att jämföras med ljudet från TV-högtalarna. I det sista testet undersöks det hur tydligt och hur högt användaren talar. I inställningarna finner man två alternativa startfraser att välja mellan: ”Hi TV” och ”Smart TV”. Den fras som är förkryssad skall uttalas då man vill inleda röststyrningen. (Jordan 2012c) Den används med andra ord för att fånga TV:ns uppmärksamhet och minskar då risken för att vanligt prat skall uppfattas som kommandon (Katzmaier 2012). I röststyrningsmenyn kan man välja vilket språk man vill ge kommandon på. Då man sagt aktiveringsfrasen så dyker det upp en rad med möjliga kommandon nere på TV-skärmen som bara är att läsa av. Om TV:n inte reagerar på kommandon som man ger den eller om den reagerar för lätt så att den gör något man inte menade, så finns det i menyn en möjlighet att ändra på känsligheten. (Jordan 2012c)



*Figur 13. Då man uttalat röststyrningens aktiveringsfras dyker möjliga kommandon upp på skärmen (van Ballegoie 2012).*

För att öka på chanserna att bli förstådd så skall man tala högt och tydligt i en så tyst omgivning som möjligt. Det rekommenderas att man uttalar kommandon med en ljudnivå som ligger vid 60 dB eller högre. Då man inleder röststyrningen med aktiveringsfrasen så sänks TV:ns volym automatiskt till fem i sådana fall då volymen är högre, så att kommandon lättare skall kunna uppfattas. (Samsungs e-Manual 2012) Då Jordan (2012c) testade röststyrningen i en av videorna i serien ”Keep It Simple” som SPSN har framställt, så var han ibland tvungen att upprepa ett kommando eller höja rösten en aning trots att han stod i ett tyst rum på ett ganska nära avstånd från TV:n. I synnerhet

då det finns en pågående diskussion eller annat ljud i bakgrunden så kan TV:n ha svårt att uppfatta kommandon. I sådana fall lönar det sig att i stället använda mikrofonen som finns på fjärrkontrollen, eftersom man då kan tala rakt i mikrofonen efter att man tryckt in knappen "voice". (Samsungs e-Manual 2012)

### 4.3 Ansiktsigenkänning

Ansiktsigenkänning kan användas då man vill komma in på Smart Hub med sitt Samsung-konto bara genom att ställa sig framför kameran. Detta kräver förstås att man först registrerar sin uppsyn på kontot. Om man väljer att använda denna funktion så sparar TV:n användarens ansiktsbild i miniatyrformat, så att den kan användas till identifiering. Genom att använda ansiktsigenkänningsfunktionen så behöver man inte slå in sitt lösenord eller sitt användarnamn. Att logga in genom att använda ansiktsigenkänning är dock inte en säker inloggningsmetod. (Samsungs e-Manual 2012) Ett bevis på detta framgick i en av "Keep It Simple"-videorna av SPSN. Där lyckades man logga in på en användares konto genom att endast visa ett fotografi av denna person för kameran. Där bevisades det även att personer som ser lika ut kan uppfattas som skilda personer, vilket innebär att tvillingar kan ha egna konton. Testet utfördes så att man på en och samma gång visade flera likadana ansiktsfotografier för kameran. (Jordan 2012b)



Figur 14. Ansiktsigenkänningsfunktionen sparar en miniatyrbild av en användares ansikte.

## 5 TESTER OCH RESULTATREDOVISNING

Detta kapitel är indelat i fyra delkapitel. I första delkapitlet behandlas rörelsestyrningen i Kinect och Samsungs Smart TV, följande delkapitel tar upp röststyrningen och det tredje delkapitlet innehåller undersökningar om ansiktsigenkänningsfunktionen. Delka-

pitlen är indelade i avsnitt som behandlar olika egenskaper av den NUI-funktion (NUI är en förkortning av natural user interface) som kapitlet handlar om. Avsnitten är uppbyggda så att det till först finns relaterade tester och konstateranden, varefter egna tester och resultatredovisning presenteras. Det sista delkapitlet består av en sammanfattad resultatredovisning.

Testerna är utförda i en miljö som följer rekommendationerna i användarmanualerna för respektive apparat. Smart TV:n testades på ett avstånd på 2 meter från TV:n och Kinect testades 1,8 meter från sensorn. Kinect placerades ovanför TV:n så att den befann sig 1,3 meter från golvet. Illuminansen i rummet mättes med en digital luxmätare, TES-1332. P.g.a. det svartspräckliga golvet och de mörka möblerna i vardagsrummet där testerna utfördes, adderades extra ljuskällor för att uppnå en tillräcklig belysning. En digital decibelmätare, AZ-8928, användes till att anpassa omgivningen så att rekommendationerna för röststyrningsfunktionerna efterföljdes. Belysningen i rummet låg mellan 100 och 185 lux, och det kontrollerades att bakgrundsljudet i rummet inte översteg 40 dB och att röststyrningskommandona uttalades med en ljudstyrka på minst 60 dB. Ifall ljus- och ljudstyrkan har avvikit från rekommendationerna i något av testfallen så anges det skilt för sig i den motsvarande testredovisningen.



Figur 15. Mätinstrumenten som användes i samband med testerna.

## 5.1 Rörelsestyrningsfunktionen i allmänhet

Katzmaier (2012) skriver att rörelsestyrningen i Samsungs Smart TV kräver mycket ljus för att fungera och att skuggor ställer till problem. Han säger dock att han ofta hade svårt att få rörelsestyrningen att fungera även om rummet var bra upplyst, och att styrningen var trög och osmidig då han lyckades aktivera funktionen. Även Mossberg (2012) hade svårigheter att få TV:n att uppfatta hans vinkningar så att han överhuvudtaget kunde starta rörelsestyrningen. Katzmaier skriver också att han ofta var tvungen att ”klicka” på ett objekt flera gånger genom att stänga och öppna handen innan han blev uppfattad, att det var hopplöst att skrolla en sida, att systemet överlag reagerade trögt och att det var frustrerande att använda rörelsestyrningsfunktionen.

**TEST 1, Kinect:** Då Kinects rörelsestyrningsfunktion testades så upplevdes styrningen vara lätthanterbar och navigeringsobjektet flyttades ganska smidigt omkring på skärmen. Styrningsfunktionen var även lagom känslig vilket innebar att man kunde utföra snabba rörelser inom ett brett område. Att välja ett objekt var dock inte det lättaste då man var tvungen att hålla handen stilla på ett ställe tills förloppsindikatorn blev full. Under testets gång påträffades inga problem med att aktivera rörelsestyrningsfunktionen. Man kunde också konstatera att det går att spela Kinect i ett mörkt rum där den enda ljuskällan är TV-skärmen. Rörelsestyrningen fungerade utan problem även vid starkt ljus.

**TEST 2, Smart TV:** Rörelsestyrningsfunktionen i Samsungs Smart TV upplevdes inte vara någon höjdare. Då man försökte klicka på ett objekt så lyckades det inte alltid vid första försöket, och om man klickade i snabb takt, t.ex. för att ändra ljudvolymen steg för steg så avbröts rörelsestyrningsfunktionen. Efter att ha testat funktionen en stund så blev man t.o.m. irriterad och kom på sig själv med att sitta och sucka högt. Det krävdes inte varje gång en vinkning för att aktivera funktionen, utan det räckte med att man höjde handflatan, vilket resulterade i att funktionen ibland aktiverades då man inte hade som avsikt att använda den. Enligt användarmanualen så borde rörelsestyrningen inte fungera på det sättet. Rörelsestyrningsfunktionen upplevdes vara alltför känslig, då man var tvungen att göra mycket små och långsamma rörelser för att det hela skulle fungera på ett nöjaktigt sätt. Ljuset påverkade rörelsestyrningen. Vid för litet ljus var det svårare

att aktivera funktionen och att klicka på ett objekt lyckades inte alltid. Navigeringspilen blev vid svagare ljus mindre känslig och flyttade sig hackigare. Om man lade till något rött i bakgrunden för att uppnå kontrast så började rörelsestyrningen fungera ganska normalt igen trots det svaga ljuset. Vid för starkt ljus så fungerade rörelsestyrningen väldigt dåligt även om man hade röd bakgrund.

**RESULTAT:** Rörelsestyrningen i Kinect får i det stora hela väl godkänt. Man kände att man hade kontroll över det hela och lärde sig snabbt hur man skulle använda funktionen. Rörelsestyrningen i Smart TV:n fungerade däremot inte på ett önskvärt sätt och man kunde t.o.m. bli frustrerad över dess tröghet. För att öka chanserna att lyckas med rörelsestyrningen i Smart TV:n så var man tvungen att anpassa sig alltför mycket, vilket ledde till att den naturliga känslan i det hela försvann.

### 5.1.1 Menynavigering med rörelser

Enligt Rivington (2010) så kunde man under ett Kinect-testtillfälle på TechRadar konstatera att det är både tungt och svårt att styra handikonen till rätt ställe och att sedan hålla den stilla där i de ca två sekunder som det krävs för att välja ett objekt. Rivington tillägger att man ansåg att rörelsestyrning med svepmetoden är enklare (d.v.s. det att man drar handen från vänster till höger eller tvärtom).

**TEST 1, Kinect:** Under testtillfället gick det bra att navigera i Xbox-menyn med handrörelser. Det är dock stor skillnad på hur bra navigeringen fungerar från spel till spel. I Dance Central fungerade rörelsestyrningen väldigt bra, i Kinect Adventures fungerade den bra eller nöjaktigt, medan spelet Kung Fu Panda 2 var svårt att navigera. I menyn i Kung Fu Panda 2 väljer man mellan de olika alternativen genom att föra handen uppåt eller neråt från den ursprungliga positionen där man håller handen höjd framför sig. Det hände att man hoppade över ett eller flera av alternativen i menyn oberoende av hur långsamma och små rörelser man än försökte göra. Rörelsestyrning med sveprörelser testades både i Xbox-menyn och i Dance Central-spelen och resultatet var väldigt positivt. Man kunde föra armen i en bred vågrät rörelse och det fungerade även då man endast gjorde ett svep med handen. Navigeringen med sveprörelser lyckades vid varje test, och det fungerade även ofta fastän man utförde vårdslösare sveprörelser.

**TEST 2, Smart TV:** Peknavigeringen i Samsungs Smart TV var litet av en utmaning. Det kändes som om man inte hade full kontroll över styrningen, och pilen kunde hoppa från ett ställe till ett annat vid snabbare rörelser. Kommandot ”tillbaka”, d.v.s. en cirkelrörelse i motsols riktning, fungerade överraskande bra och det lyckades de flesta gångerna. Funktionen för att skrolla, där man ”drar” t.ex. en internetsida uppåt eller neråt, upplevdes som värdelös eftersom den inte fungerade nästan alls. Då man använde webbrowsern så hände det även att peknavigeringspilen på skärmen stannade upp.

**RESULTAT:** Att navigera i menyer med Kinects rörelsestyrningsfunktion fungerade i de flesta fallen bra. Stora skillnader påträffades dock då man testade navigeringen i olika spel. Man tappade lusten att spela spel såsom Kung Fu Panda 2, eftersom navigeringen fungerade så dåligt. Användningen av sveprörelser upplevdes vara en navigeringsmetod som fungerar bra. Att navigera i Smart TV:ns menyer med rörelser gick inget vidare. Vissa saker fungerade nöjaktigt, medan andra fungerade dåligt.

### 5.1.2 Rörelsestyrningens responstid

I en stor del av Kinect-spelen kan man lägga märke till en fördröjning som uppstår efter det att man utfört en rörelse innan den animerade figuren på skärmen utför motsvarande rörelse. Fördröjningens omfattning påverkas även av rörelsens hastighet, vilket innebär att fördröjningen inte är lika märkbar då man utför långsamma rörelser. (Rivington 2010)

**TEST 1, Kinect:** Det tog ca en till två sekunder att aktivera rörelsestyrningsfunktionen i Xbox-menyn, och då kördes ansiktsgenkänningsfunktionen på samma gång så att rätt användarprofil kunde loggas in. Överlag så var rörelsestyrningens responstid bra, under en sekund, och Kinect hängde bra med i de rörelser man gjorde. En liten eftersläpning kunde nog märkas, men även här så var det stor skillnad från spel till spel. I Xbox-menyn var responstiden ca 0,5 sekunder och i Dance Central var reaktionen väldigt god. Fördröjningen märktes tydligare desto snabbare rörelser man utförde. I t.ex. Kinect Adventures så reagerade man på fördröjningen och kunde uppleva att den var irriterande, då man t.ex. hoppade i luften och nästan var tillbaka på golvet innan figuren i spelet

hann påbörja hoppet. Även i Kung Fu Panda 2 var fördröjningen så tydlig att den kunde upplevas som störande.

**TEST 2, Smart TV:** De gånger då det gick att aktivera rörelsestyrningsfunktionen genom att endast hålla upp handen så tog det ca 0,5 sekunder innan man fick kontroll över pil-ikonen. Att klicka på ett objekt tog lika länge. Fördröjningen märktes, men den var ändå så liten att man inte blev irriterad.

**RESULTAT:** Kinects och Smart TV:ns rörelsestyrningsfunktioner hade båda bra responstid i de inbyggda menyerna. Vissa Kinect-spel hade dock en märkbar fördröjning, och det kan hända att fallet även är så i vissa Smart TV-applikationer, men detta undersöktes inte under dessa tester.

### 5.1.3 Rörelsestyrningens påfrestning

Katzmaier (2012) skriver att han tyckte det var tungt att använda rörelsekontrollen i Samsungs Smart TV eftersom man är tvungen att hålla armen i luften.

**TEST 1, Kinect:** Långvarig menynavigering var utmattande för armen om man inte sänkte armen emellanåt. Man kunde dock styra handikonen genom att hålla sin hand i vilken riktning som helst, och att låta handen hänga så att handflatan var riktad mot golvet var betydligt mindre ansträngande än att t.ex. hålla handflatan riktad mot TV:n.

**TEST 2, Smart TV:** För att testa rörelsestyrningens fysiska påfrestning så spelade man Angry Birds på TV:n. Man blev nästan direkt trött i armen och hann inte skjuta iväg många fåglar innan det började kännas obehagligt så att man fick lust att sänka armen. Det var verkligen tungt att hålla upp handen en längre tid. Man var tvungen att hålla handflatan riktad mot TV:n hela tiden, eftersom rörelsestyrningsfunktionen avbryts om systemet inte ser en handflata. Det var mycket tyngre än att hålla handflatan riktad mot golvet såsom man kunde göra då man använde Kinect.

**RESULTAT:** Både Kinects och Smart TV:ns rörelsestyrningsfunktioner är tunga att använda i sådana fall då man är tvungen att hålla armen i luften en litet längre stund

utan att sänka den emellanåt. Smart TV:n är dock tyngre att styra än Kinect, som styrs med olika delar av kroppen och som även är mer flexibel vad gäller handrörelser.

## 5.2 Röststyrningen i allmänhet

Enligt Mossberg (2012) så fungerar inte röststyrningen i Samsungs Smart TV något vidare. Han skriver att han hade problem med att kommunicera med apparaten eftersom den inte reagerade på hans kommandon eller så uppfattade den kommandona fel och att han ofta måste upprepa en kommandofras några gånger innan någonting hände. Mossberg säger att dessa problem uppstod trots att han testade apparaten i en ljudlös omgivning på ett sådant avstånd från TV:n som det anges i användarmanualen. Katzmaier (2012) skriver att kanalbyte genom att säga kanalnumret dit man vill komma inte fungerade nästan alls. Vidare så säger han att Samsungs Smart TV inte klarade av att tolka kommandon rätt då de gavs i ett snabbt tempo. Han klassar inte röststyrningen i Samsungs Smart TV så högt eftersom han upplever att fjärrkontrollen fungerar bättre och är lättare att använda. Katzmaier tycker överlag att röststyrningsfunktionen i Samsungs Smart TV är ganska begränsad.

Följande test utfördes i tyst omgivning där även TV-volymen var avstängd.

**TEST 1, Kinect:** Röststyrningen fungerade riktigt bra, bortsett från att det fanns några kommandon som systemet hade svårt att uppfatta och som krävde upprepning. Man kunde konstatera att Kinect kräver normalt taltempo eftersom det vid för raskt eller för långsamt prat inte hände någonting. Röststyrningen fungerade även då man talade med lägre volym som låg kring 47 dB. Vissa ord fungerade bättre än andra då man talade lågt, och det fanns även sådana kommandon som fungerade då man viskade (t.ex. "Xbox" och "social"). Kommandona "home" och "settings" var problematiska, eftersom det ofta inte hände någonting fastän man uttalade dem med en klar och tydlig röst. Det hela kunde sluta med att man försökte ändra på sitt uttal eller nästan skrek åt systemet i ett försök att bli uppfattad. Det hände även att man upprepade samma kommando om och om på samma sätt och att man sedan plötsligt blev uppfattad. Om man uttalade "home" med ett utdraget "h" så hände det oftare att man blev uppfattad, men



det kändes dock onaturligt och var ansträngande. Att tala med en mörk basröst fungerade utan problem, men om man talade i falsett så spelade själva uttalet en större roll.

**TEST 2, Smart TV:** Röststyrningen fungerade riktigt bra. Vid kanalbyte påträffades märkliga feltolkningar såsom att kanal nio uppfattades som kanal tre och att kanal sju kunde tolkas som kanal fyra eller fem. Kanal 17 tolkades nästan konstant som kanal 28. Vid snabbt prat blev man missförstådd eller så blev man ombedd att upprepa kommandot. Långsamt prat fungerade ibland, medan systemet andra gånger hann tolka ett kommando som något annat kommando innan man hunnit tala till punkt. Att ändra tonläge förorsakade inga större problem förutom att en mörk basröst krävde att man talade med en högre ljudstyrka.

**RESULTAT:** Både Kinects och Smart TV:ns röststyrning fungerade ganska bra i tyst omgivning. Det fanns vissa kommandon som Kinect hade svårt att uppfatta, men systemet gjorde nästan aldrig feltolkningar. Smart TV:n gjorde däremot nog feltolkningar alltid då och då, speciellt vid kommandon som innehöll en siffra.

### 5.2.1 Röststyrning och omgivningens ljudnivå

Katzmaier (2012) skriver att det inte alltid var så lätt att bli uppfattad då han använde den inbyggda mikrofonen på Samsungs Smart TV, och att t.o.m. lite oljud i omgivningen ställde till problem. Han upplevde att resultatet blev bättre då han använde fjärrkontrollens mikrofon.

Hur omgivningens ljudnivå påverkar röststyrningen i Kinect och Samsungs Smart TV testades på två olika sätt. Under det första testet hade man en radio på i bakgrunden. På det stället där testpersonen befann sig låg bakgrundsljudet då mellan 40 och 65 dB. Inför det andra testet så stängdes radion av och så höjde man i stället TV:ns ljudvolym så mycket att den som mest var 67 dB, för att kunna testa hur bra röststyrningen fungerar då det omgivande ljudet kommer från en källa som ligger nära apparaternas mikrofoner.

**TEST 1, Kinect:** Om man talade med normal ljudstyrka så fungerade vissa kommandon såsom ”social”, medan andra ord endast uppfattades om man höjde rösten så att

man överröstade radion. ”Apps”, ”home” och ”settings” är exempel på sådana kommandon som var speciellt svåra att få fram över bakgrundsljudet. Systemet kunde alltså gå minste om något kommando, men det gjorde å andra sidan sällan feltolkningar av det man sade. Om man höjde rösten så att man talade litet högre än normalt så fungerade röststyrningen riktigt bra trots bakgrundsljudet. Röststyrningen fungerade bra även då TV:ns ljudvolym aktiverades. Man testade till sist att höja TV:ns ljudvolym ganska mycket från 67 dB, och man kunde då konstatera att Kinects röststyrning fungerar ganska bra trots högt ljud från TV:n. Om man inte blev uppfattad så behövde man endast höja på rösten så att man överröstade TV:ns egna ljud.

**TEST 2, Smart TV:** Röststyrningen fungerade ganska bra om man talade med normal ljudstyrka då man hade på radion, förutom då man skulle öppna någon applikation. Om man t.ex. sade ”Smart Hub” så hände det att systemet öppnade programmen ”search all”, ”e-manual”, ”all share” eller ”camera”, men inte det man var ute efter. Vid bakgrundsljud var problemet då alltså mera det att TV:n gjorde feltolkningar än det att den inte hörde att man sade någonting. Röststyrningens aktiveringskommando ”Hi TV” fungerade nästan alltid oberoende av om man talade med normal volym eller litet svagare. Då man aktiverar röststyrningen så sänks TV:n ljudvolym automatiskt om den är för hög. För att då kunna testa hur röststyrningsfunktionen klarar av TV-ljudvolym som går upp till 67 dB, så var man tvungen att höja volymen först efter att man aktiverat röststyrningsläget. Vid ljud från TV:n så fungerade röststyrningen nöjaktigt då man talade normalt till litet högre. En del feltolkningar påträffades och så var man ibland tvungen att höja rösten. Det hände även att TV:n började lyssna på sig själv och trodde att TV-ljudet innehöll kommandon. Det slutade då med att den öppnade någon applikation såsom Smart Hub eller Samsung Apps eller så bad den att man skulle upprepa kommandot eller frågade om man befann sig i en bullrig miljö.

**RESULTAT:** Lika som då man testade rörelsestyrningsfunktionerna i tyst omgivning så kunde man även i det här fallet konstatera att Smart TV:ns problem var det att den kunde göra feltolkningar av det som sades, medan Kinect sällan gjorde det. Kinect fungerade ganska bra trots ljud i omgivningen, och de gånger den hade problem med att uppfatta kommandon så löste man det hela genom att höja rösten en aning. Smart TV:ns röststyrningsfunktion klarade bättre av bakgrundsljud från en radio än den klarade av

ljud från själva TV:n. Då det fanns ljud i omgivningen så lyckades röststyrningen bättre om man använde fjärrkontrollens mikrofon.

### 5.2.2 Röststyrningens responstid

Enligt Schramm (2012) så kan man i Kinect-spelet Mass Effect 3 lägga märke till en ca en sekunds fördröjning efter att man sagt ett kommando till att operationen utförs i spelet. Han skriver att fördröjningen beror på det att Kinect först måste uppfatta att man uttalat ett kommando och sedan räkna ut vad kommandot betyder innan systemet utför en åtgärd. Schramm påpekar att det antagligen går lika snabbt, om inte t.o.m. snabbare att använda kontrollen till att styra händelser i spelet.

**TEST 1, Kinect:** Under testtillfället kunde man konstatera att det tar ca en sekund eller litet längre för någonting att hända efter att man uttalat ett kommando. Fördröjningen är trots allt så liten, så man störs egentligen inte av den.

**TEST 2, Smart TV:** Lika som då man använde Kinects rörelsestyrning, så tog det ungefär en sekund eller litet längre innan Smart TV:n gjorde någonting efter att man hade sagt ett kommando. Att slå på TV:n med kommandot "Hi TV starta" tog längre. Det räckte över en och en halv sekund innan TV:n kom igång efter att hela kommandot hade uttalats. Det kanske inte låter som en så mycket längre tid, men här lade man nog märke till fördröjningen och fick lust att upprepa kommandot då man trodde att TV:n inte hade hört att man sagt något. Det tog allt som allt ca 2,8 sekunder innan TV:n slogs på från det att man började tala.

**RESULTAT:** Rörelsestyrningens responstid bör vara så kort som möjligt så att man inte störs av fördröjningen. Responstiden är en av de faktorer som avgör om det lönar sig att använda röststyrningsfunktionen eller inte. Om det tar för lång tid innan något sker så kommer man antagligen hellre att använda kontrollen. Både Kinects och Smart TV:ns rörelsestyrningsfunktioner hade överlag acceptabel responstid. Smart TV:ns kommando "Hi TV starta" hade dock för lång responstid.

### 5.2.3 Röststyrning och uttal

Enligt Schramm (2012) så fungerar Kinects röststyrningsfunktion även om man talar med olika accenter.

**TEST 1, Kinect:** Kinects röststyrningsfunktion testades med engelska som användarspråk. Testet bevisade att röststyrningen fungerar oberoende av om man talar amerikansk eller brittisk engelska. Vissa kommandon fungerade även om man endast uttalade en del av ordet, såsom "Xbo" i stället för "Xbox" och "soc" i stället för "social".

**TEST 2, Smart TV:** Smart TV:ns röststyrningsfunktion testades med svenska som användarspråk. Vid kanalbyte eller volymändring, då man gav kommandot "kanal sju" eller "sju", så uppfattades sjuan så när som varje gång som 20. Man kunde till slut konstatera att det krävdes ett rikssvenskt uttal för att bli förstådd. Om man sade "kanal 17" på finlandssvenska så tolkade TV:n det som 28. Bortsett från problemen som uppstod då man uttalade vissa siffror så fungerade röststyrningen överlag oberoende av om man talade finlandssvenska, rikssvenska eller t.o.m. sabbodialekt. Röststyrningen fungerade även om man uttalade orden såsom de skrivs, exempelvis då man sade "sänk ljudet" i stället för "senk judet" och "källa" i stället för "tschella". Smart TV:ns röststyrning fungerade också fastän man inte uttalade hela ordet, d.v.s. man kunde exempelvis säga "Hi T" i stället för "Hi TV" och "Search a" i stället för "search all".

**RESULTAT:** Kinects och Smart TV:ns röststyrningsfunktioner var ganska flexibla vad gäller uttalet av kommandon. Smart TV:ns röststyrning krävde dock ett rikssvenskt uttal vid vissa kommandon, vilket framstod som väldigt irriterande då man som finlandssvensk försökte kommunicera med TV:n.

## 5.3 Ansiktsigenkänningen i allmänhet

Mossberg (2012) skriver att han under ett testtillfälle inte fick ansiktsigenkänningsfunktionen i Samsungs Smart TV att fungera, trots att han t.o.m. flyttade på sig så han stod nära kameran, rakt framför TV:n.

**TEST 1, Kinect:** Kinects ansiktsigenkänningsfunktion fungerar överlag riktigt bra. Någon gång kunde det hända att identifieringen misslyckades trots att man varken hade gjort några radikala utseendemässiga förändringar eller ändrat på ljuset i rummet, men om man gjorde ett nytt försök så brukade inloggningen sedan nog lyckas. Efter att systemet har tagit referensbilder så att ansiktsigenkänningsfunktionen för en användarprofil kan aktiveras, så meddelar Kinect att det lönar sig att komma tillbaka och ta nya bilder ifall man dramatiskt ändrar på utseendet såsom att man börjar använda glasögon.

**TEST 2, Smart TV:** Så länge som man hade tillräckligt med ljus i omgivningen och inte ändrade på det egna utseendet eller ljuset i rummet så lyckades ansiktsidentifieringen för det mesta. Vid testtillfället hände det ibland att själva ansiktsigenkänningsfunktionen fick något fel. Det kunde vara frågan om att kameran endast visade svart eller att miniatyrbilder av testpersonen irrade omkring på skärmen. Smart TV:n tar ingen referensbild innan den ser att det verkligen finns ett ansikte inom kamerans synfält, vilket kunde konstateras då TV:n lät bli att ta ett foto av testpersonen då denna höll en röd pingisracket framför ansiktet.

**RESULTAT:** I situationer där både användarens utseende och miljön där användaren befinner sig i inte har genomgått några större förändringar så fungerar Kinects och Smart TV:ns ansiktsidentifieringsfunktioner riktigt bra. Smart TV:n kräver dock att rummet är ordentligt upplyst.

### 5.3.1 Hur omgivningens belysning påverkar ansiktsigenkänningen

För att Kinects ansiktsigenkänningsfunktion skall fungera optimalt, så krävs det en viss belysning i rummet. Man får hjälp till att anpassa belysningen enligt rekommendationerna genom att följa instruktionerna i apparatens ”Kinect Tuner”. Direkt solljus rekommenderas inte, varken på användaren eller rakt mot kameran. Ett av solen bländat ansikte gör ansiktsdragen otydligare för kameran och i sådana fall då solljus lyser upp en användare bakifrån så kan det hända att ansiktet skuggas. (Frum 2010)

Inför dessa tester tog man först en inloggningsbild av testpersonen i ”normalt” ljus. Därefter höjdes först belysningen så att den hade ett maxvärde på 190 lux och då prö-

vade testpersonen om apparaterna klarade av att identifiera hennes ansikte. Efter det prövade man ansiktsgenkänningsfunktionen på motsvarande sätt, men denna gång genom att sänka ljuset så att det låg mellan 50 och 70 lux.

**TEST 1, Kinect:** Större ljusförändringar ställde till med problem för Kinects ansiktsgenkänningsfunktion. Om man dock tog nya bilder av användaren i det nya ljusförhållandet så lyckades inloggningen igen.

**TEST 2, Smart TV:** Testerna bevisade att Smart TV:ns ansiktsgenkänningsfunktion är känslig för ljusförändringar. Då man hade höjt på ljuset så misslyckades inloggningen. Vid det sänkta ljusförhållandet var det inte heller lätt att få ansiktsgenkänningsfunktionen att fungera, men vid testtillfället så lyckades det trots allt vid tredje inloggningsförsöket. Man prövade även att ta en ny inloggningsbild av testpersonen i den ljusare omgivningen och då lyckades inloggningen efter det, men då man gjorde på motsvarande sätt för den mörkare omgivningen så kunde testpersonen fortfarande inte logga in.

**RESULTAT:** Både Kinects och Smart TV:ns ansiktsgenkänningsfunktioner är känsliga för ljusförändringar, vilket minskar på den naturliga känslan då man använder funktionen. Om belysningen i rummet förändras lite mer så bör användaren ta nya referensbilder av sig själv. Smart TV:ns ansiktsgenkänningsfunktion kommer dock inte att fungera trots nya referensbilder i sådana fall då det är för mörkt i omgivningen.

### **5.3.2 Ansiktsgenkänningsfunktionens flexibilitet**

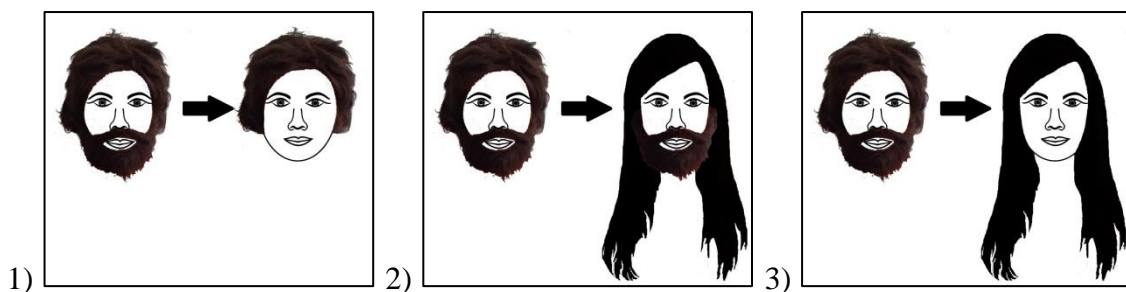
Jordan (2012b) bevisade att ansiktsgenkänningsfunktionen i Samsungs Smart TV är flexibel, eftersom han kunde logga in sig även efter att han limmat en stor mustasch i ansiktet.

Hur flexibla Kinects och Smart TV:ns ansiktsgenkänningsfunktioner sist och slutligen är har testats på följande sätt.

**Ansiktsbehåring och olika hårstilar:** I den första delen av detta test använde testpersonen en peruk för att täcka in sitt långa hår så att det blev som kortklippt. Testpersonen

fick dessutom klä sig i skägg plus mustasch. Därefter sparade man referensbilder av personen i apparaternas minne. Det tas alltid en bild för Smart TV:ns ansiktsigenkänningsfunktion och flera bilder i Kinects fall, eftersom det där tas bilder av användaren då han eller hon befinner sig på olika ställen i rummet. Till del två av detta test tog man referensbilder av användaren där hon såg ut som normalt med håret hängande löst över axlarna.

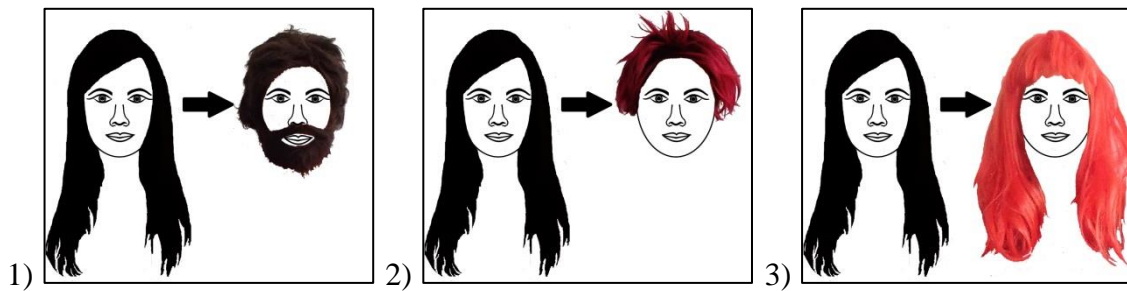
**TEST 1, Kinect:** I den första delen av testet så lyckades testpersonen logga in med ansiktsigenkänningsfunktionen efter att hon hade tagit av sig skägget plus mustaschen. Inloggningen lyckades även om hon i stället tog av sig peruken och visade sitt egna långa hår. Då hon lämnade bort både peruken och skägget plus mustaschen så kände systemet däremot inte längre igen henne. I den andra delen av testet så kunde testpersonen inte logga in då hon var klädd i skägg plus mustasch. Om hon satte upp håret i en hästsvans så kunde hon nog logga in. Då testpersonen använde en peruk med kort spretigt hår som inte täckte pannan så lyckades också ansiktsigenkänningen, men då hon klädde sig i en peruk som bestod av rött långt hår med pannlugg så lyckades inloggningen bara så länge som pannluggen inte rörde ögonbrynen.



*Figur 16. Ansiktsbehåring och olika hårstilar. Första delen av testet och dess delmoment.*

**TEST 2, Smart TV:** I den första delen av testet så gick det inte längre att logga in efter att skägget plus mustaschen hade tagits bort. Samma sak hände de flesta gångerna då man i stället tog bort peruken och lät håret "växa". I den andra delen av testet så lyckades inte identifieringen efter att man lagt till skägg plus mustasch. Att Jordan (2012b) kunde logga in efter att han limmat en mustasch i ansiktet berodde antagligen på att den inte täckte lika mycket av ansiktet som skägg plus mustasch helheten i detta test gjorde. Om man satte upp håret i en hästsvans eller använde en peruk med kort spretigt hår som

inte täckte pannan, så gick det däremot nog att logga in. Ansiktsigenkänningen lyckades även då testpersonen hade på sig en peruk som bestod av rött långt hår plus pannlugg.



Figur 17. Ansiktsbehåring och olika hårstilar. Andra delen av testet och dess delmoment (hästsvans saknas).

**Olika ansiktsuttryck:** Inför dessa tester sparade man referensbilder i Kinect och Smart TV:n där testpersonen hade en vanlig uttryckslös min.

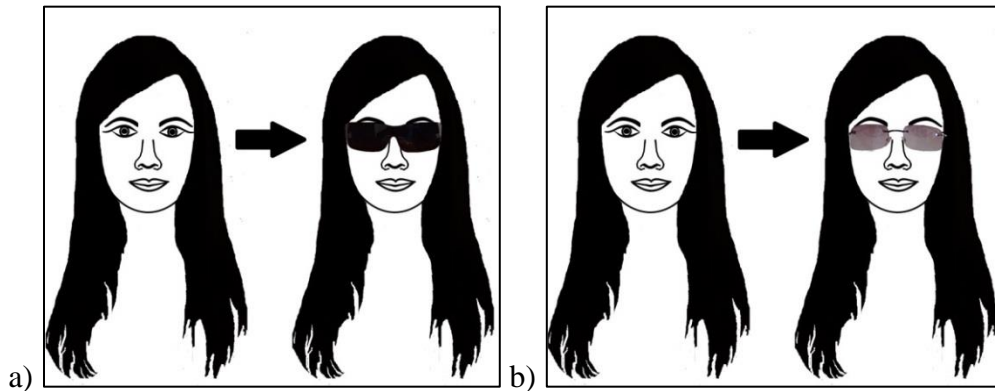
**TEST 1, Kinect:** Testpersonen provade till att börja med att visa tänderna. Då hon log ordentligt så att tänderna syntes bra så misslyckades identifieringen. Så länge som testpersonen endast log så att en liten del av tänderna syntes, och därmed inte använde kindmusklerna så mycket att stora delar av ansiktet ändrade utseende, så lyckades däremot inloggningen. Inga problem påträffades i sådana situationer då testpersonen såg arg ut och därmed rynkade ögonbrynen kraftigt, och inte heller då hon såg förvånad ut och gapade samt höjde ögonbrynen ordentligt.

**TEST 2, Smart TV:** Då testpersonen visade tänderna så lyckades ansiktsigenkänningen så länge som hon endast visade en liten del av tänderna. Då hon log med hela munnen så misslyckades inloggningen. Samma sak hände då hon såg arg ut. Inloggningen misslyckades även då testpersonen såg förvånad ut genom att hon både lyfte på ögonbrynen och gapade. Om hon dock endast gjorde endera, d.v.s. antingen lyfte på ögonbrynen eller gapade, så lyckades identifieringen så länge som ansiktsuttrycket i sin helhet inte förändrades alltför mycket.

**Användning av glasögon och huvudbonad:** Inför dessa tester tog man referensbilder av testpersonen där hon varken hade glasögon eller någon huvudbonad på sig.

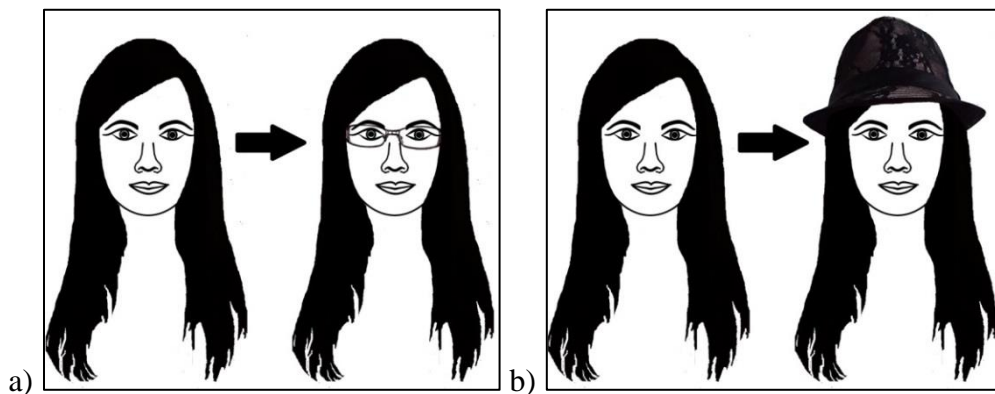


**TEST 1, Kinect:** Testpersonen lyckades inte logga in då hon hade mörka solglasögon med litet större lins på sig. Då hon däremot bar smala, ganska genomskinliga solglasögon så lyckades inloggningen. Då testpersonen använde glasögon så lyckades hälften av inloggningsförsöken och hälften inte. Vid inloggningsförsöken där testpersonen hade någonslags huvudbonad på sig så kände Kinect igen henne så länge som huvudbonaden inte drogs så långt ner över pannan att den skymde ögonbrynen.



*Figur 18. Användning av solglasögon. Hur klarar ansiktigenkänningsfunktionen av  
a) mörka solglasögon b) ljusa solglasögon med mindre lins?*

**TEST 2, Smart TV:** Ansiktigenkänningen misslyckades varje gång då testpersonen hade glasögon eller solglasögon på sig, oberoende av om de hade stora eller små linser. Testpersonen hade hela tiden samma ansiktsuttryck och det fanns ingen lampa som bländade glasögonlinserna. Då testpersonen prövade olika huvudbonader, så lyckades inloggningsförsöken så länge som huvudbonaderna, eller skuggan av dem, inte täckte för mycket av pannan. För mycket skulle ungefär innebära att de rör ögonbrynen.

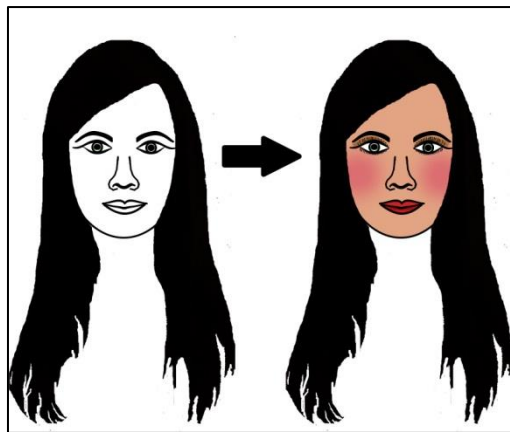


*Figur 19. Användning av glasögon eller huvudbonad. Hur klarar ansiktigenkänningsfunktionen av  
a) glasögon b) en hatt?*

**Användning av smink:** Inför detta test tog man referensbilder av testpersonen då hon var osminkad.

**TEST 1, Kinect:** Ansiktsigenkänningen lyckades utan problem oberoende av hur mycket testpersonen sminkade sig. Först testade man vanligt dagssmink och ökade därefter på sminket tills testpersonen bar väldigt stark blåfärgad ögonskugga, tjock kajal, rött läppstift, mycket foundation och rouge. Trots detta väldigt överdrivna smink kände Kinect igen användaren.

**TEST 2: Smart TV:** Då testpersonen sminkades med vanligt dagssmink så lyckades ansiktsigenkänningen utan problem. Inloggningen lyckades även efter att testpersonen blivit sminkad med starkare ögonsmink, gräsligt tjock foundation och färggrant läppstift. Först då testpersonen sminkades med extra stark blåfärgad ögonskugga, tjock kajal, klarrött läppstift, mycket foundation och stark rouge så misslyckades identifieringen. Sminket var dock vid det laget så tjockt och överdrivet att ingen normalt skulle sminka sig på det sättet.



*Figur 20. Man testade hur smink påverkar Kinects och Smart TV:ns ansiktsigenkänningsfunktioner.*

**RESULTAT:** I det stora hela så var Smart TV:ns ansiktsigenkänningsfunktion mindre flexibel jämfört med Kinects ansiktsidentifieringsfunktion. Om man bytte hårstil eller frisyr, eller började använda huvudbonad så ledde det i de flesta fallen inte till problem för apparaterna, så länge som pannan inte täcktes för mycket. Smart TV:n klarade inte

av att identifiera en användare om denna gjorde sig av med sitt eventuella skägg och sin mustasch efter att referensbilden hade tagits, eller tvärtom om ett skägg och en mustasch hade kommit till. Kinects ansiktsidentifiering misslyckades endast i det senare fallet. Smart TV:ns ansiktsigenkänningsfunktion var känsligare än Kinects ansiktsidentifieringsfunktion också då det gällde olika miner. Under testtillfället då en användare hade satt solglasögon eller glasögon på sig efter att referensbilden hade tagits, så kände Smart TV:n inte igen användaren längre. Kinect borde nog klara av att känna igen en person med glasögon så länge som glasögonen inte har för stor eller för mörk lins. Varken Kinect eller Smart TV:n hade svårigheter att identifiera en användare som använde smink.

### **5.3.3 Ansiktsigenkänning och personer med olika hudfärg**

En artikel på websidan GameSpot ledde till misstankar om att Kinects ansiktsigenkänningsfunktion inte klarar av att hantera mörkhyade personer. Till följd av denna artikel så undersöktes saken noggrannare av konsumenter och den slutliga domen som framgick ur konsumentrapporter visade att det trots allt bara var frågan om ljusproblem i rummet. Enligt en talesman på Microsoft så fungerar apparaten för alla oberoende av hudfärg. (Frum 2010) Efter detta utförde man på GameSpot fler tester. Nu lyckades identifieringen av de mörkhyade personer som tidigare misslyckats. Då en av dessa personer bytte skjorta från en ljusblå till en svart under samma ljusförhållanden så misslyckades dock igenkänningen, likaså då ytterligare en mörkhyad person försökte bli identifierad. I båda fallen kalibrerade man apparaten några gånger men identifieringen misslyckades trots det. (Sinclair 2010)

För att testa hur hudfärgen inverkar på ansiktsigenkänningen tog man först en inloggningsbild av testpersonen såsom hon såg ut normalt, d.v.s. där ansiktet var ganska blekt. Därefter målades denna person med maskeradfärg så att ansiktet och halsen såg solbrända ut, så att man sedan kunde testa om inloggningen fungerar efter att man har fått färg i ansiktet. I den andra delen av testet målades testpersonens ansikte och hals mörkbruna. Testpersonen hade även svarta vantar på händerna för att täcka all ljus hudfärg och på huvudet bar hon en svart afro-peruk. Slutresultatet var att hon liknade en afrikan.

Därefter tog man en referensbild av testpersonen i denna utstyrsel, så att hon sedan kunde göra inloggningsförsök som mörkhyad.



*Figur 21. Testpersonen blev maskerad som mörkhyad.*

**TEST 1, Kinect:** Kinects ansiktsigenkänningsfunktion fungerade utan problem då testpersonen var “solbränd”. Inga inloggningsproblem påträffades heller då testpersonen var maskerad som en mörkhyad. Hon kunde logga in även om hon var svartklädd från topp till tå och också efter att man dessutom hade sänkt belysningen i rummet.

**TEST 2, Smart TV:** Då testpersonen hade blivit “solbränd” efter att den senaste referensbilden hade tagits så lyckades inte längre ansiktsigenkänningen. Det intressanta är dock det att identifieringen misslyckades även i den andra delen av testet då testpersonen var maskerad som mörkhyad. Detta hände trots att den sparade bilden i minnet ju exakt motsvarade det som kameran såg eftersom den hade tagits precis innan. Inloggningen misslyckades fastän testpersonen hade suttit blickstillä utan att ändra en min sedan referensbilden hade tagits. Upprepade inloggningsförsök gjordes. Man testade även att ändra ljuset så att ansiktet blev bättre upplyst och att lägga till röd kontrastfärg i bakgrunden, men det hjälpte inte heller saken.

**RESULTAT:** Kinect verkar fungera oberoende av hudfärg. Om det händer att identifieringen misslyckas så beror det antagligen på någon annan faktor än hudfärgen. Såsom Frum (2010) skriver så kan det eventuellt bero på för svag belysning i rummet, men under testtillfällena inför detta examensarbete så påträffades dock inga sådana problem.

Smart TV:ns ansiktsgenkänningsfunktion kräver nya referensbilder då man blivit solbränd efter att den senaste referensbilden är tagen. Det aningen rasistiska är att Smart TV:n inte klarar av att identifiera mörkhyade personer trots att referensbilden är en exakt kopia av det som kameran ser vid inloggningstillfället.

## 5.4 Resultatredovisning

Kinects rörelsestyrningsfunktion fungerar så bra i vissa spel att man mycket gärna använder den. Funktionen kan alltså klassas som användbar. Det finns dock sådana Kinect-spel där man har satsat väldigt dåligt på rörelsestyrningen, vilket gör att man önskar att man kunde styra spelet med en kontroll i stället. Menynavigering kan vara ett område där man hellre använder kontrollen eftersom det i vissa fall kan vara den snabbare lösningen. Fastän rörelsestyrningen inte fungerar perfekt så finns det spel såsom dansspelet Dance Central, där rörelsestyrning faktiskt är den bästa lösningen. Man kan då röra sig fritt utan att behöva hålla i en kontroll och Kinect följer dessutom med rörelser från hela kroppen. Smart TV:ns rörelsestyrning är däremot väldigt bristfällig och trög att använda, och man vill ju inte använda en funktion som man blir nervös på. Funktionen var så jobbig att använda att man inte ens fann intresse att använda den då TV:n var ny och man testade apparaten för första gången. Såsom Katzmaier (2012) skriver så är rörelsestyrningsfunktionen i Samsungs Smart TV alltså inte användbar.

I det stora hela så går det snabbare att använda en fjärrkontroll än att använda röststyrningen i Kinect och Samsungs Smart TV. Varför skulle man då vilja tala med sin TV eller Kinect? I Kinects fall där man spelar utan en kontroll så kan röststyrningen vara användbar jämsides med rörelsestyrningsfunktionen. Att använda röststyrning till att slå på Smart TV:n kan vara användbart, t.ex. då man kommer in i rummet och inte genast kommer åt fjärrkontrollen. I ett vanligt vardagsrum där det ofta finns ljud i omgivningen så lönar det sig att använda mikrofonen på fjärrkontrollen i stället för mikrofonen på TV:n. Lika som Katzmaier (2012) påpekar så kan man dock undra varför man då inte lika bra skulle kunna använda knapparna på fjärrkontrollen till att styra TV:n, då man redan har fjärrkontrollen i handen. Apparaterna har en viss flexibilitet när det gäller uttalet av kommandon, och då man använder röststyrningen så vill man ju också tala så

fritt som möjligt. Man vill i varje fall inte t.ex. som finlandssvensk bli tvungen att tala rikssvenska för att bli förstådd såsom Smart TV:n krävde vid vissa kommandon.

Varken Kinects eller Smart TV:ns ansiktsigenkänningsfunktioner fungerar problemfritt. Ibland känner de inte igen en användare fastän man tycker att de borde göra det. Speciellt dåligt är det faktum att systemen inte känner igen en användare bara för att han eller hon gör vissa miner. Kinects ansiktsigenkänningsfunktion fungerar bättre än Smart TV:ns identifieringsfunktion. Båda är dock nog användbara så länge som man inte har för mörk hudfärg, eftersom Smart TV:n inte verkar klarar av att känna igen mörkhyade personer. Då identifieringen lyckas så slipper man mycket enklare undan än om man t.ex. måst slå in användarnamn och lösenord på ett tangentbord som finns på skärmen.

Som Mossberg (2012) och Katzmaier (2012) skriver så kan det konstateras att Samsungs Smart TV:s NUI-funktioner kräver en hel del förbättringar, men att Samsung nog är inne på rätt spår. Så som systemet fungerar just nu så använder man nog hellre fjärrkontrollen. Rörelsestyrningsfunktionen skulle kunna lämnas bort helt och hållet. Om man förbättrar systemet så kan det naturliga användargränssnittet i TV:n bli mer användbart. När det gäller Kinect så är kontrollen bra att ha som stöd t.ex. vid menynavigering. Kinects röststyrning är som en bonusfunktion, d.v.s. användbar men man klarar sig bra utan den. Både i Kinect och i Smart TV:n så använder man antagligen nog ansiktsigenkänningsfunktionen som första alternativ och loggar in manuellt bara ifall identifieringen misslyckas.

## **6 SLUTSATSER OCH DISKUSSION**

Är då dagens naturliga användargränssnitt verkligen till någon nytta? Då Jordan (2012a) testade Samsungs Smart TV så sades det att det naturliga användargränssnittet gör användningen roligare. Det är kanske just det som det handlar om ännu i det här läget. Ett naturligt användargränssnitt tillåter en användare att styra en apparat på nymodiga sätt, vilket kan kännas spännande. För att funktionerna skall ha något värde i det långa loppet så måste de däremot vara flexibla och fungera på ett så problemfritt och naturligt sätt som möjligt. Det kan annars hända att man tappat intresset för dem och övergår till att använda en fjärrkontroll för att det är enklare. Den fysiska påfrestningen vid rörelse-

styrning är också en orsak till att man kanske föredrar fjärrkontrollen. Interaktionsmetoden bör inte heller ta längre eller vara svårare att använda än knappar på en fjärrkontroll för att man skall välja att använda den.

Avslutningsvis så kan jag säga att jag anser att fysisk kontakt, med andra ord tillgång till vanliga knappar och pekskärmar, nog fortfarande behövs då man kommunicerar med maskiner. Utvecklingen måste ännu gå mycket framåt innan naturliga användargränssnitt börjar fungera så bra att man skulle kunna börja fundera på att slopa fjärrkontrollen. Det finns i dagens läge nog sådana ansiktsgenkänningsfunktioner, röststyrningsfunktioner och rörelsestyrningsfunktioner som är riktigt användbara, men de fungerar då bra tillsammans med de klassiska interaktionsmetoderna, inte ensamma. Även fastän man skulle förbättra Smart TV:ns naturliga användargränssnitt avsevärt så ser jag dock inte poängen med att kommunicera med en TV genom att göra teckenspråk i luften. Fallet är helt annat då det gäller Kinect, eftersom hela idén där är att styra spel med kropps-rörelser. Jag anser att nya interaktionsmetoder verkligen skall bidra med något användbart, eftersom ingen i längden orkar använda en funktion som inte är enkel och effektiv att använda. Fördelarna med röststyrning kan vara många bara systemet är uppbyggt på ett bra sätt. Som exempel kan nämnas att röststyrning av en telefon kan höja säkerheten vid bilkörning. Likaså är en bra ansiktsgenkänningsfunktion till nytta, eftersom den underlättar inloggningar och liknande.

## KÄLLOR

- Ban, Kyu-Dae; Lee, Jaeyeon; Hwang, Dae Hwan & Chung, Yun-Koo. 2008, Face Image registration methods using Normalized Cross Correlation, *International Conference on Control, Automation and Systems (ICCAS) 2008*. Konferenspublikation utgiven av IEEE, s. 2408-2411, oktober 2008.
- Benzon. 2012, Samsung 46" ES8005 (LED) – forbruger review, BZN Reviews. Tillgänglig: <http://bzn.dk/reviews/samsung-46-es8005-led-forbruger-review/> Hämtad 8.4.2013.
- Bladon, Anthony. 1994, Speech recognition applications should focus on the human factors, *Telemarketing. Tidskrift utgiven av Technology Marketing Corporation*, vol. 13, nr 5, s. 65-68, november 1994.
- Blake, Joshua. 2012, *Natural User Interfaces in .NET, WPF 4, Surface 2, and Kinect*, Early Access Edition, Manning Publications Co., 47 s.
- Cao, Zhimin; Yin, Qi; Tang, Xiaou & Sun, Jian. 2010, Face recognition with learning-based descriptor, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Konferenspublikation utgiven av IEEE, s. 2707-2714, augusti 2010.
- Cruz, Leandro; Lucio, Djalma & Velho, Luiz. 2012, Kinect and RGBD Images: Challenges and Applications, *2012 25<sup>th</sup> SIBGRAPI Conference on Graphics, Patterns and Images Tutorials*. Konferenspublikation utgiven av IEEE, s. 36-49, augusti 2012.
- Fook, C. Y.; Hariharan, M.; Yaacob, Sazali & AH, Adom. 2012, A Review: Malay Speech Recognition and Audio Visual Speech Recognition, *International Conference on Biomedical Engineering (ICoBE)*. Konferenspublikation utgiven av IEEE, s. 479-484, februari 2012.
- Forsythe, Thomas & Green, Marvin. 2012, Advancements in 3-D sensing technology implemented by the Kinect, *University of Pittsburgh Swanson School of Engineering, Conference Session: C7, paper # 2338*. 6 s., mars 2012.
- Frum, Larry. 2010, Lightning affects Kinect's face recognition, report says, *CNN Tech*. Tillgänglig: [http://articles.cnn.com/2010-11-04/tech/kinect.dark.skin\\_1\\_facial-recognition-skin-tones-lighting?\\_s=PM:TECH](http://articles.cnn.com/2010-11-04/tech/kinect.dark.skin_1_facial-recognition-skin-tones-lighting?_s=PM:TECH) Hämtad 11.12.2012.
- Gantenbein, Douglas. 2011, Helping Kinect Recognize Faces, *Microsoft Research*. Tillgänglig: <http://research.microsoft.com/en-us/news/features/kinectfacereco-103111.aspx> Hämtad 7.11.2012.
- How Microsoft's PrimeSense-based Kinect really works. 2011, *Electronic Design. Tidskrift utgiven av Penton Media*, vol. 59, nr 5, s. 28-30, juli 2011.






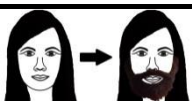

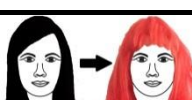
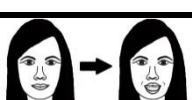
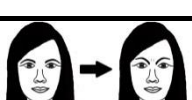
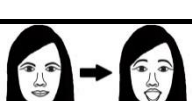
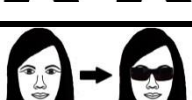

- Ionescu, Dan; Ionescu, Bogdan; Gadea, Cristian & Islam, Shahidul. 2011, An Intelligent Gesture Interface for Controlling TV Sets and Set-Top Boxes, *2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*. Konferenspublikation utgiven av IEEE, s. 159-164, maj 2011.
- Jordan, Myles. 2012a, Keep It Simple, 2012 Smart TV: Gesture Control, *Samsung Product Support Network (SPSN)*. Tillgänglig: [http://support-us.samsung.com/spsn/specialized.jsp?keyword=2012 smart tv&video id=1697146187001](http://support-us.samsung.com/spsn/specialized.jsp?keyword=2012%20smart%20tv&video_id=1697146187001) Hämtad 16.12.2012.
- Jordan, Myles. 2012b, Keep It Simple, 2012 Smart TV: Face Recognition, *Samsung Product Support Network (SPSN)*. Tillgänglig: [http://support-us.samsung.com/spsn/specialized.jsp?keyword=2012 smart tv&video id=1697161381001](http://support-us.samsung.com/spsn/specialized.jsp?keyword=2012%20smart%20tv&video_id=1697161381001) Hämtad 16.12.2012.
- Jordan, Myles. 2012c, Keep It Simple, 2012 Smart TV: Voice Control, *Samsung Product Support Network (SPSN)*. Tillgänglig: [http://support-us.samsung.com/spsn/specialized.jsp?keyword=2012 smart tv&video id=1697024016001](http://support-us.samsung.com/spsn/specialized.jsp?keyword=2012%20smart%20tv&video_id=1697024016001) Hämtad 16.12.2012.
- Katzmaier, David. 2012, Samsung Smart Interaction: Hands-on with voice and gesture control, *CNET*. Tillgänglig: [http://reviews.cnet.com/8301-33199\\_7-57411497-221/samsung-smart-interaction-hands-on-with-voice-and-gesture-control/](http://reviews.cnet.com/8301-33199_7-57411497-221/samsung-smart-interaction-hands-on-with-voice-and-gesture-control/) Hämtad 16.12.2012.
- Khoshelham, K. 2011, Accuracy analysis of Kinect depth data, *ISPRS Workshop Laser Scanning 2011. Konferensdokumentation utgiven av International Society of Photogrammetry and Remote Sensing (ISPRS)*, vol. 38, s. 133-138, augusti 2011.
- Kramer, Jeff; Burrus, Nicolas; Echtler, Florian; Herrera C., Daniel & Parker, Matt. 2012, *Hacking the Kinect*, New York, Apress, 251 s.
- Knies, Rob. 2011a, Kinect Audio: Preparedness Pays Off, *Microsoft Research*. Tillgänglig: <http://research.microsoft.com/en-us/news/features/kinectaudio-041311.aspx> Hämtad 13.11.2012.
- Knies, Rob. 2011b, Kinect Body Tracking Reaps Renown, *Microsoft Research*. Tillgänglig: <http://research.microsoft.com/en-us/news/features/kinectskeletal-092711.aspx> Hämtad 29.1.2013.
- Leyvand, Tommer; Meekhof, Casey; Wei, Yi-Chen; Sun, Jian & Guo, Baining. 2011, Kinect Identity: Technology and Experience, *Computer. Tidskrift utgiven av the IEEE Computer Society*, vol. 44, nr 4, s. 94-96, april 2011.
- Lone, Manzoor Ahmad; Zakariya S. M. & Ali Rashid. 2011, Automatic Face Recognition System by Combining Four Individual Algorithms, *International Conference on Computational Intelligence and Communication Networks (CICN) 2011. Konferenspublikation utgiven av IEEE*, s. 222-226, oktober 2011.

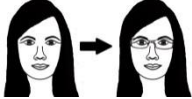

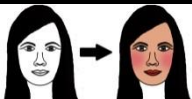
- Mejia, Paola. 2010, Joystiq Labs: Kinect a new experience, *Joystiq*. Tillgänglig: <http://en.wikinoticia.com/Technology/games/66203-joystiq-labs-kinect-a-new-experience> Hämtad 8.4.2013.
- Michel, Theo. 2011, Introducing ... the Kinect Tuner!, *123Kinect*. Tillgänglig: <http://123kinect.com/kinect-forums/thread-1012.html> Hämtad 8.4.2013.
- Moni, M.A. & Ali, A.B.M.S. 2009, HMM based hand gesture recognition: A review on techniques and approaches, *2nd IEEE International Conference on Computer Science and Information Technology 2009 (ICCSIT)*. Konferenspublikation utgiven av IEEE, s. 433-437, augusti 2009.
- Mossberg, Wlaler S. 2012, Samsung's Smart TV Isn't As Smart As It Thinks It Is, *Wall Street Journal*. Tidning utgiven av Dow Jones & Company Inc, augusti 2012.
- Naone, Erica. 2011, Microsoft Kinect, How the device can respond to your voice and gestures, *Technology Review*. Tidskrift utgiven av Massachusetts Institute of Technology, s. 82-83, januari 2011.
- Panwar, Meenakshi. 2012, Hand gesture recognition based on shape parameters, *International Conference on Computing, Communication and Applications (ICCCA) 2012*. Konferenspublikation utgiven av IEEE, s. 1-6, februari 2012.
- Rai, Joel. 2012, Remote decontrol, *Business Today*. Tidskrift utgiven av Living Media India Limited, juli 2012.
- Rankin, Bob. 2007, Voice recognition, *Ask Bob Rankin*. Tillgänglig: [http://askbobrankin.com/voice\\_recognition.html](http://askbobrankin.com/voice_recognition.html) Hämtad 8.4.2013.
- Rivington, James. 2010, Microsoft Kinect for Xbox 360 review - You become the Xbox 360 controller with Microsoft's motion peripheral, *TechRadar*. Tillgänglig: <http://www.techradar.com/reviews/gaming/games-consoles/controllers/microsoft-kinect-for-xbox-360-905010/review> Hämtad 13.12.2012.
- Samsungs e-Manual (för 46" Smart 3D LED-TV ES8005). 2012. Tillgänglig: [http://downloadcenter.samsung.com/content/UM/201203/20120327100751967/18\\_EPDVBEUE\\_SWE.pdf](http://downloadcenter.samsung.com/content/UM/201203/20120327100751967/18_EPDVBEUE_SWE.pdf) Hämtad 5.2.13.
- Schramm, Mike. 2012, Testing Kinect in Mass Effect 3 (plus, demo coming February 14) [update: more demo details], *Joystiq*. Tillgänglig: <http://www.joystiq.com/2012/01/18/testing-kinect-in-mass-effect-3-plus-demo-coming-february-14/> Hämtad 13.12.2012.
- Shotton, Jamie; Fitzgibbon, Andrew; Cook, Mat; Sharp, Toby; Finocchio, Mark; Moore, Richard; Kipman, Alex & Blake, Andrew. 2011, Real-Time Human Pose Recognition In Parts From Single Depth Images, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2011*. Konferenspublikation utgiven av IEEE, s. 1297-1304, juni 2011.

- Sinclair, Brendan. 2010, Kinect has problems recognizing dark-skinned users?, *GameSpot*. Tillgänglig: <http://www.gamespot.com/news/kinect-has-problems-recognizing-dark-skinned-users-6283514> Hämtad 11.12.2012.
- Sung, Kelvin. 2011, Recent Videogame Console Technologies, *Computer. Tidsskrift utgiven av IEEE Computer Society*, s. 91-93, februari 2011.
- Tashev, Ivan. 2012, Audio for Kinect: pushing it to the limit, *CREST Symposium on Human-Harmonized Information Technology, Kyoto University*. Tillgänglig: <http://www.ar.media.kyoto-u.ac.jp/crest/sympo12/program/S4-Tashev.pdf> Hämtad 14.3.2013.
- van Ballegoie, Eric. 2012, Samsung ES8000 reviewed: the next level of Smart, *Hardware.info*. Tillgänglig: <http://uk.hardware.info/reviews/2722/5/samsung-es8000-reviewed-the-next-level-of-smart-hi-tv!> Hämtad 8.4.2013.
- Zhang, Zhengyou. 2012, Microsoft Kinect Sensor and Its Effect, *IEEE MultiMedia. Tidsskrift utgiven av the IEEE Computer Society*, vol. 19, nr 2, s. 4-10, februari 2012.

## BILAGOR

Tabell 2. Ansiktsigenkännings flexibilitet.

	KINECT			SMART TV		
	Lyckades ansiktsigenkänningen?			Lyckades ansiktsigenkänningen?		
	Ja	Ibland	Nej	Ja	Ibland	Nej
	X					X
	X				X	
			X			X
			X			X
	X			X		
	X			X		
			X			X
	X					X
	X				X	
			X			X
	X					X

		X				X
	X			X		
	X			X		